Amino acid change in an Orchid desaturase enables mimicry of the pollinator's sex pheromone

Sedeek, Khalid E M ; Whittle, Edward ; Guthörl, Daniela ; Grossniklaus, Ueli ; Shanklin, John ; Schlüter, Philipp M

Abstract: Mimicry illustrates the power of selection to produce phenotypic convergence in biology [1]. A striking example is the imitation of female insects by plants that are pollinated by sexual deception of males of the same insect species [2-4]. This involves mimicry of visual, tactile, and chemical signals of females [2-7], especially their sex pheromones [8-11]. The Mediterranean orchid Ophrys exaltata employs chemical mimicry of cuticular hydrocarbons, particularly the 7-alkenes, in an insect sex pheromone to attract and elicit mating behavior in its pollinators, males of the cellophane bee Colletes cunicularius [11-13]. A difference in alkene double-bond positions is responsible for reproductive isolation between O. exaltata and closely related species, such as O. sphegodes [13-16]. We show that these 7-alkenes are likely determined by the action of the stearoyl-acyl-carrier-protein desaturase (SAD) homolog SAD5. After gene duplication, changes in subcellular localization relative to the ancestral housekeeping desaturase may have allowed proto-SAD5's reaction products to undergo further biosynthesis to both 7- and 9-alkenes. Such ancestral coproduction of two alkene classes may have led to pollinator-mediated deleterious pleiotropy. Despite possible evolutionary intermediates with reduced activity, amino acid changes at the bottom of the substrate-binding cavity have conferred enzyme specificity for 7-alkene biosynthesis by preventing the binding of longer-chained fatty acid (FA) precursors by the enzyme. This change in desaturase function enabled the orchid to perfect its chemical mimicry of pollinator sex pheromones by escape from deleterious pleiotropy, supporting a role of pleiotropy in determining the possible trajectories of adaptive evolution.

DOI: https://doi.org/10.1016/j.cub.2016.04.018

Amino acid change in an orchid desaturase enables mimicry of the pollinator's sex pheromone

Khalid E. M. Sedeek[1], Edward Whittle[2], Daniela Guthörl[3], Ueli Grossniklaus[3], John Shanklin[2], and Philipp M. Schlüter[1,*]

[1]Department of Systematic and Evolutionary Botany and Zurich-Basel Plant Science Centre, University of Zurich, Zollikerstr. 107, CH-8008 Zurich, Switzerland

[2]Department of Biology, Brookhaven National Laboratory, 50 Bell Ave., Upton, NY 11973, USA

[3]Department of Plant and Microbial Biology and Zurich-Basel Plant Science Centre, University of Zurich, Zollikerstr. 107, CH-8008 Zurich, Switzerland

[*]Correspondence: *philipp.schlueter@systbot.uzh.ch*

SUMMARY

Mimicry illustrates the power of selection to produce phenotypic convergence in biology [1]. A striking example is the imitation of female insects by plants that are pollinated by sexual deception of males of the same insect species [2-4]. This involves mimicry of visual, tactile and chemical signals of females [2-7], especially their sex pheromones [8-11]. The Mediterranean orchid *Ophrys exaltata* employs chemical mimicry of cuticular hydrocarbons, particularly the 7-alkenes, in an insect sex pheromone to attract and elicit mating behaviour in its pollinators, males of the cellophane bee *Colletes cunicularius* [11-13]. A difference in alkene double-bond positions is responsible for reproductive isolation between *O. exaltata* and closely related species, such as *O. sphegodes* [13-16]. We show that these 7-alkenes are likely determined by the action of the stearoyl-acyl-carrier-protein desaturase (SAD) homologue SAD5. After gene duplication, changes in subcellular localization relative to the ancestral housekeeping desaturase may have allowed proto-SAD5's reaction products to undergo further biosynthesis to both 7- and 9-alkenes. Such ancestral co-production of two alkene classes may have led to pollinator-mediated deleterious pleiotropy. Despite possible evolutionary intermediates with reduced activity, amino acid changes at the bottom of the substrate binding cavity have conferred enzyme specificity for 7-alkene biosynthesis by preventing the binding of longer-chained fatty acid (FA) precursors by the enzyme. This change in desaturase function enabled the orchid to perfect its chemical mimicry of pollinator sex pheromones by escape from deleterious pleiotropy, supporting a role of pleiotropy in determining the possible trajectories of adaptive evolution.

RESULTS AND DISCUSSION

*SAD*s Are Expressed In Flower Labellum Epidermis

Chemical mimicry of sex pheromones by *Ophrys* orchids is mediated by cuticular hydrocarbons that are likely derived from very-long-chain FA biosynthesis and deposited on the flower surface as wax layer components [16-18]. The flower epidermis is therefore expected to express the genes controlling pollinator specificity via alkene double-bond position: *SAD2* for (*Z*)-9- and (*Z*)-12-alkenes common in *O. sphegodes* MILLER [16], and potentially other members of the three known clades of orchid *SADs* (*SAD1–SAD6*; *SAD2*-clade: *SAD1/2*; *SAD3*-clade: only *SAD3*; *SAD5*-clade: *SAD4/5/6* [15]), some postulated to be responsible for (*Z*)-7-alkenes abundant in *O. exaltata* subsp. *archipelagi* (GÖLZ & H.R.REINHARD) DEL PRETE (hereafter, *O. exaltata*) [16, 18]. This hypothesis was tested by RNA *in situ* hybridization (ISH) on sections of flower labellum, the modified petal with which pollinators interact. ISH on thin sections of labella from greenhouse-grown *O. exaltata* and *O. sphegodes* for members of all three *SAD* clades revealed epidermal expression (Figs. 1, S1). This is consistent with previous PCR-based expression-level data [15, 16], but revealed no difference in spatial expression pattern between the different desaturase homologues in labella. Therefore, involvement in alkene synthesis cannot be ruled out for any orchid *SAD* based upon their spatial expression patterns.

SADs Show Diversification in Enzyme Function

Mimicry of the *C. cunicularius* female sex pheromone by *O. exaltata* crucially depends on the specific synthesis of 7-alkenes by the orchid; by themselves (even without a flower), these 7-alkenes are physiologically detectable by *C. cunicularius* males and sufficient to elicit the behavioural responses required for successful orchid pollination [11-13]. Enzymatic activities of recombinant *Ophrys* SAD3, SAD4 and SAD5 encoded by putatively functional alleles were investigated using *in vitro* assays and coupled gas chromatography (GC)-mass spectrometry (MS). SAD1 and SAD2 had previously been characterized [16], SAD4 had no detectable activity, and characterization of SAD6 was infeasible (no soluble expression). SAD3 catalyses the ubiquitous acyl-carrier-protein (ACP)-linked desaturation reaction [19-23] of stearate (a C_{18} FA with zero double bonds, denoted 18:0) to oleate (with one double bond at position 9, denoted $18{:}1\Delta^9$), and of 16:0-ACP to $16{:}1\Delta^9$-ACP (Fig. 2). This mirrors the function of archetypal desaturases [22-24] and is consistent with SAD3's hypothesized role as a housekeeping desaturase, previously proposed given its ubiquitous expression and lack of statistical

association with floral alkene accumulation [16]. By contrast, SAD5-A (the putatively functional allelic isoform of SAD5 [15]) is a specialized 16:0-ACP Δ^9 desaturase that has very low activity on 18:0-ACP substrate (Fig. 2; Table S1) and shows remarkable specificity towards 16:0-ACP (~35× more specific than SAD3). Elongation and decarbonylation of its product 16:1Δ^9-ACP, with its double bond 7 carbons from the aliphatic end, are expected to result in the production of 7-alkenes (Fig. 2I). Therefore, SAD5 catalyses a reaction that explains the strong statistical association of its gene expression with 7-alkene production during floral development and in natural orchid populations [15]. This enzyme is likely responsible for alkene-based mimicry of *C. cunicularius* sex pheromones in *O. exaltata*. Like SAD2, which catalyses the formation of precursors for 9- and 12-alkenes in *O. sphegodes* [16], SAD5's enzymatic function overlaps with that of SAD3, implying a potential for functional non-independence (epistasis) among SADs. This also raises the question of why SAD3 does not substantially contribute to alkene biosynthesis in *Ophrys* flowers. The finding that SAD3 is not strongly associated with 7- and 9-alkene production [15], despite being capable of producing the relevant precursors, suggests that its reaction products are channelled into different downstream metabolic pathways.

Altered Subcellular Localization of SAD5 May Direct its Reaction Products into a Different Pathway

Altered protein-protein interactions of SADs may be responsible for different metabolic fates of desaturase reaction products. Plant SAD proteins and *de novo* FA synthesis up to C_{18} generally occur in plastids [17, 19, 22, 25]. Since bioinformatic predictions of N-terminal transit peptides (TPs; putatively responsible for plastid localization) were inconclusive (Table S2), the subcellular localization of SAD1–SAD5 (but not SAD6; see Experimental Procedures) was tested experimentally. *SAD* coding sequences were fused to green fluorescent protein (GFP) and subcellular localization of SAD-GFP assessed by confocal microscopy in intact *Nicotiana* leaf protoplasts. All proteins were broadly localized to regions showing chlorophyll autofluorescence, suggesting plastid association (Fig. S2). However, the distribution of GFP signal differed between SAD paralogues: SAD1 and SAD3 showed a diffuse signal, whereas SAD2, SAD4 and SAD5 displayed a distinctively speckled (punctate) pattern (Figs. 1, S2), consistent with localization to plastoglobules (lipoprotein subcompartments of plastids) [cf. 26, 27, 28]. The difference in subcellular localization between the putative housekeeping (SAD3)

and alkene precursor-producing enzymes (SAD2/SAD5) highlights the possibility of spatial separation of these processes and channelling of metabolites to different destinations. Comparison between different functional regions within the proteins based on sequence polymorphism data of representative *SAD* alleles (Supplemental Experimental Procedures) showed that putative TP ($p < 10^{-4}$; Fisher's exact test; potentially affected by poor alignability) and ACP-interacting region ($p = 0.019$; Table S3) harbour more amino acid changes than expected. This implies that SAD2, SAD3 and SAD5 may differ in their binding affinity for acyl thioesters of the different *Ophrys* ACP isoforms [29]. Overall, differences in ACP interaction and subcellular localization suggest that potential metabolite-mediated epistasis between SAD3 and SAD2/SAD5 may be prevented by different metabolic fates of their reaction products. Potential epistasis between SAD2 and SAD5 is unlikely to occur *in planta* as species-specific gene expression was observed, with *SAD2* only highly expressed in *O. sphegodes* and *SAD5* in *O. exaltata* [15, 16].

Evolution of Protein Structure and Function Suggests a Role for Deleterious Pleiotropy in Adaptation

Given our molecular data and the known evolutionary relationships of *SAD* paralogues [15], the ancestor of the *SAD5*-clade is inferred to have featured a plastoglobule-like subcellular localization (Fig. 3). Likewise, the ancestral enzymatic activity of orchid SADs is inferred as 16:0 Δ^9 and 18:0 Δ^9 desaturation, like SAD3 and other, well-characterized housekeeping desaturases [19-24, 30]. Therefore, the ancestor of SAD5 would likely have pleiotropically produced precursors for 7- and 9-alkenes, and it appears plausible that these may have been directed to the very-long-chain FA pathway, resulting in floral alkenes. *C. cunicularius* males prefer 7-alkene blends without co-occurring 9-alkenes [11]. Given the likely recent time-scale of *O. exaltata* evolution [31-33], this preference may well have been the selective force [see 15, 33] that favoured loss of 18:0-ACP Δ^9 desaturase activity in the evolutionary lineage leading to modern SAD5-A, thereby allowing escape from deleterious pleiotropy. Our protein structural modelling suggests that this happened by loss of effective C_{18} substrate binding, specifically due to amino acid changes at the bottom of the substrate binding cavity preventing the efficient insertion of a C_{18} acyl chain into the protein by steric hindrance (Fig. 3B): In SAD5, three amino acids obstruct the lower end of the substrate pocket (V141, I206, and particularly, F145). This leaves enough space to accommodate the shorter C_{16} acyl chain, with carbon 9

positioned near the active site, to allow the production of 7-alkene precursors. To test the simplest hypothesis that these three residues in SAD5 alone may be responsible for its increased specificity towards 16:0-ACP, mutated SAD5-A variants were engineered reverting these amino acids to their inferred ancestral states. All combinations of ancestral/present-day amino acids were generated and assayed (Table S1, Fig. S3). The resurrected ancestral triple-replacement protein (t_{-3} in Fig. 4) retained considerable activity on 16:0-ACP (about a third that of SAD5-A), while activity on 18:0-ACP was restored to levels similar to SAD3, supporting the functional importance of the tested residues (along with single/double site replacements at positions 145/206; Table S1). However, functional non-independence (epistasis) of the three residues was evident. In particular, enzymatic activity was nearly abolished in all single/double (but not triple) mutants reverting V141 to M141 (Fig. S3) – a segregating polymorphism in present-day *O. exaltata* populations [15] (V in SAD5-A and M in SAD5-B allelic variants). Taken together, these results indicate that pollinator-mediated selection against 9-alkenes may have favoured key amino acid changes in an ancestor of SAD5 that abolished its activity on C_{18}, thereby enabling specific 7-alkene mimicry in *Ophrys* orchids without the negative pleiotropy associated with the ancestral coproduction of 7- and 9-alkenes. Minimization of deleterious pleiotropy is an important explanatory factor for understanding the molecular mechanisms of evolutionary transitions in another pollination-relevant trait, flower colour [34]. The case of SAD5 is therefore consistent with the idea that pleiotropic constraints are generally important for determining the trajectory of evolution realized at the molecular level.

The Evolutionary Trajectory of Modern SAD5 May Have Contained Loss-of-Function Variants

The onset of selection on SAD5 by *C. cunicularius* remains unclear because the historical pollinator environment is unknown. However, the finding that some evolutionary intermediates *en route* to SAD5 (Fig. 4, before M→V141) were effectively loss-of-function (LOF) variants reveals a potential twist to the evolutionary history of modern SAD5: Since typical *Ophrys* pollinators should not be able to exert selection on a desaturase in the absence of alkene production, LOF raises the question of how present-day SAD5 could have evolved. At least three non-mutually exclusive scenarios could explain this paradox: (i) unevaluated epistasis among amino acids, (ii) population genetic processes, and (iii) biochemical properties of desaturase dimers. First, consistent with non-additive effects among the experimentally tested SAD5 residues, proteins may frequently experience epistasis between sites [35-

39]. Since we only evaluated three amino acids in SAD5's substrate binding pocket experimentally, while ignoring 27 changes elsewhere in the protein, other, compensating mutations may have permitted SAD5's ancestors to retain activity despite seemingly LOF configurations. Second, our *in vitro* activity data are expected to reflect the situation in homozygous plants. However, a LOF point mutation would initially have been present in a heterozygous state. Large effective population sizes and standing genetic variation in these orchids [13, 32] imply a reduced effect of genetic drift. Without a clear fitness disadvantage in the heterozygous state, a LOF allele may therefore have persisted in the population for long enough (probably at low frequency) to accumulate the mutation (M→V141) that restored activity to the previously defunct allele (Fig. 4). Evolutionary modelling suggests that once a functional *SAD5-A* allele is formed, pollinator-mediated selection would rapidly increase its allele frequency [33]. Third, SAD proteins act as dimers and recent biochemical work has shown that only one desaturase subunit is active at a time, and that heterodimers with one LOF and one functional subunit have the same net activity as homodimers with two functional subunits, whereas only LOF/LOF homodimers (25%) are inactive [40]. Thus, ancestor proteins of SAD5, or their combinations in heterozygotes, may well have possessed desaturase activity sufficient for alkene production and associated pollinator attraction. This situation would have changed with the evolution of the modern allele, which may have responded to a different selection pressure (possibly by *C. cunicularius*), favouring the 7-alkenes nowadays observed in *O. exaltata*.

SAD5 Is a Speciation Gene

Chemical mimicry in *Ophrys* is an 'automatic magic trait' [*sensu* 41] because it is both under divergent selection and directly involved in reproductive isolation (RI) [13, 15, 42]. By its involvement in mimicry, *SAD5* is therefore a major 'magic gene' and, moreover, an ecological speciation gene, following Nosil's [43] criteria: (i) via the alkene profile, it affects a critical component of RI [13, 14]; (ii) it diverged prior to completion of speciation (since RI is strong but incomplete [14, 32]); (iii) it is likely under selection [15], and (iv) trait manipulation and modelling suggest that it increases RI [15, 33].

In conclusion, *SAD5* arose by gene duplication and diversification from an ancestral housekeeping desaturase. Changes in molecular interactions likely enabled its reaction products to undergo pleiotropic 7- and 9-alkene biosynthesis. Pollinator-mediated selection likely favoured loss of

activity on 18:0-ACP, resulting in specific 7-alkene production and sex pheromone mimicry in *O. exaltata*.

EXPERIMENTAL PROCEDURES

RNA *in situ* Hybridization

For each of the three *SAD* clades [15], the most divergent clade-specific ~500 bp of sequence were subcloned in sense and antisense direction, RNA probes synthesized and ISH performed on thin sections of *O. exaltata* and *O. sphegodes* labella, as described previously [44, 45] and detailed in Supplemental Experimental Procedures.

Subcellular Localization

The presence and length of a TP for plastid localization of *Ophrys* SADs was predicted using `ChloroP` 1.1 [46] (Table S2). Subcellular localization was assessed by confocal laser-scanning microscopy of transiently expressed SAD-GFP fusion proteins in *N. benthamiana* leaf protoplasts, as detailed in Supplemental Experimental Procedures. Full-length fusions of SAD1-SAD5 with GFP were produced; a separate SAD6-GFP was not produced because of nearly perfect N-terminal sequence identity with SAD5 (see Supplemental Experimental Procedures). GFP and chlorophyll images were overlaid and autotuned using `Adobe Photoshop CS5`.

SAD Functional Characterization

Different *SAD* genes were amplified from clones [15, 16] and transferred into expression vectors as detailed in Supplemental Experimental Procedures. Putative TP sequences were removed, leaving mature desaturase-encoding sequences followed by six histidine residues at the C-terminus. Where soluble expression of recombinant protein could be achieved in *Escherichia coli*, proteins were extracted, purified and used in biochemical assays of enzyme activity, using C_{16} and C_{18} substrates as described previously [16] and detailed in Supplemental Experimental Procedures. The functions of specific amino acids in SAD5 were assayed in the same way after introduction of the desired mutations (V141M, F145L and I206V), or combinations thereof, into the expression construct by a gene synthesis service (GenScript, Piscataway, NJ, USA), reverting the present-day sites to their ancestral states (see below); all possible combinations of amino acids at these positions were tested.

Protein Evolutionary Analysis

Three-dimensional models of molecular structure were obtained for all proteins used for functional analysis as described in Supplemental Experimental Procedures. Models were then visualized and compared in RasMol 2.7.5.2 [47], including a superimposed substrate-containing desaturase structure [48] to facilitate comparison of the substrate binding channel. Given the known phylogenetic relationships among gene copies [15, 16] (Fig. 3), representative *SAD* sequences (plus outgroups) were used to reconstruct ancestral sequences using PAML software [49] as detailed in Supplemental Experimental Procedures. Likewise, ancestral character states were inferred for three characters, namely C_{16} and C_{18} enzymatic activity and subcellular localization, by maximum parsimony using the R 2.15.2 [50] ape library 3.0-11 [51] (see Supplemental Experimental Procedures). Putatively functional regions based on published information were used to annotate the alignment of *SAD1-SAD6* alleles as detailed in Supplemental Experimental Procedures. The degree of synonymous versus nonsynonymous change within these regions compared to the rest of the protein was assessed using Fisher's exact test.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three tables, three figures and Supplemental Experimental Procedures and can be found with this article online at #####.

AUTHOR CONTRIBUTIONS

Conceptualization, Funding Acquisition: U.G./J.S./P.M.S.; Investigation K.E.M.S./E.W./D.G./P.M.S.; Formal Analysis: K.E.M.S/J.S./P.M.S.; Visualization, Writing – Original Draft: K.E.M.S./P.M.S.; Writing – Review & Editing: all authors; Software, Supervision, Project Administration: P.M.S.

ACKNOWLEDGEMENTS

The authors are grateful for financial support by the University of Zurich, the Swiss National Science Foundation (grants 31003A_130796 to PMS and 31003A_141245 to UG), and the Office of Basic Energy Sciences of the US Department of Energy (EW and JS). We thank Célia Baroux, Bastien Christ, Christof Eichenberger, Luzia Guyer, Stefan Hörtensteiner, Sharon Kessler, Stephen Schauer

and Anja Schmidt for discussion and technical advice, as well as Kelsey Byers, Florian Schiestl and Shuqing Xu for comments.

REFERENCES

1. Wiens, D. (1978). Mimicry in plants. Evol. Biol. *11*, 365-403.
2. Schlüter, P.M., and Schiestl, F.P. (2008). Molecular mechanisms of floral mimicry in orchids. Trends Plant Sci. *13*, 228-235.
3. Kullenberg, B. (1961). Studies in *Ophrys* pollination. Zool. Bidr. Uppsala *34*, 1-340.
4. Paulus, H.F., and Gack, C. (1990). Pollinators as prepollinating isolation factors: evolution and speciation in *Ophrys* (Orchidaceae). Israel J. Bot. *39*, 43-79.
5. Ågren, L., Kullenberg, B., and Sensenbaugh, T. (1984). Congruences in pilosity between three species of *Ophrys* (Orchidaceae) and their hymenopteran pollinators. Nova Acta Reg. Soc. Sci. Upsaliensis *Ser. V, C, 3*, 15-25.
6. Gaskett, A.C., and Herberstein, M.E. (2010). Colour mimicry and sexual deception by Tongue orchids (*Cryptostylis*). Naturwiss. *97*, 97-102.
7. Gaskett, A.C. (2014). Color and sexual deception in orchids: progress toward understanding the functions and pollinator perception of floral color. In Darwin's orchids: then and now, R. Edens-Meier and P. Bernhardt, eds. (Chicago and London: University of Chicago Press).
8. Schiestl, F.P., Peakall, R., Mant, J.G., Ibarra, F., Schulz, C., Franke, S., and Francke, W. (2003). The chemistry of sexual deception in an orchid-wasp pollination system. Science *302*, 437-438.
9. Schiestl, F.P., Ayasse, M., Paulus, H.F., Löfstedt, C., Hansson, B.S., Ibarra, F., and Francke, W. (1999). Orchid pollination by sexual swindle. Nature *399*, 421-422.
10. Bohman, B., Phillips, R.D., Menz, M.H.M., Berntsson, B.W., Flematti, G.R., Barrow, R.A., Dixon, K.W., and Peakall, R. (2014). Discovery of pyrazines as pollinator sex pheromones and orchid semiochemicals: implications for the evolution of sexual deception. New Phytol. *203*, 939-952.

11. Mant, J.G., Brändli, C., Vereecken, N.J., Schulz, C.M., Francke, W., and Schiestl, F.P. (2005). Cuticular hydrocarbons as sex pheromone of the bee *Colletes cunicularius* and the key to its mimicry by the sexually deceptive orchid, *Ophrys exaltata*. J. Chem. Ecol. *31*, 1765-1787.

12. Vereecken, N.J., and Schiestl, F.P. (2008). The evolution of imperfect floral mimicry. Proc. Natl. Acad. Sci. USA *105*, 7484-7488.

13. Mant, J.G., Peakall, R., and Schiestl, F.P. (2005). Does selection on floral odor promote differentiation among populations and species of the sexually deceptive orchid genus *Ophrys*? Evolution *59*, 1449-1463.

14. Xu, S., Schlüter, P.M., Scopece, G., Breitkopf, H., Gross, K., Cozzolino, S., and Schiestl, F.P. (2011). Floral isolation is the main reproductive barrier among closely related sexually deceptive orchids. Evolution *65*, 2606-2620.

15. Xu, S., Schlüter, P.M., Grossniklaus, U., and Schiestl, F.P. (2012). The genetic basis of pollinator adaptation in a sexually deceptive orchid. PLoS Genet. *8*, e1002889.

16. Schlüter, P.M., Xu, S., Gagliardini, V., Whittle, E.J., Shanklin, J., Grossniklaus, U., and Schiestl, F.P. (2011). Stearoyl-acyl carrier protein desaturases are associated with floral isolation in sexually deceptive orchids. Proc. Natl. Acad. Sci. USA *108*, 5696-5701.

17. Haslam, T.M., and Kunst, L. (2013). Extending the story of very-long-chain fatty acid elongation. Plant Sci. *210*, 93-107.

18. Perera, M.A.D.N., Qin, W., Yandeau-Nelson, M., Fan, L., Dixon, P., and Nikolau, B.J. (2010). Biological origins of normal-chain hydrocarbons: a pathway model based on cuticular wax analyses of maize silks. Plant J. *64*, 618-632.

19. Shanklin, J., and Cahoon, E.B. (1998). Desaturation and related modifications of fatty acids. Annu. Rev. Plant Physiol. Plant Mol. Biol. *49*, 611-641.

20. McKeon, T.A., and Stumpf, P.K. (1982). Purification and characterization of the stearoyl-acyl carrier protein desaturase and the acyl-acyl carrier protein thioesterase from maturing seeds of safflower. J. Biol. Chem. *257*, 12141-12147.

21. Lindqvist, Y., Huang, W., Schneider, G., and Shanklin, J. (1996). Crystal structure of Δ^9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. EMBO J. *15*, 4081-4092.

22. Shanklin, J., and Somerville, C.R. (1991). Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. Proc. Natl. Acad. Sci. USA *88*, 2510-2514.

23. Kachroo, A., Shanklin, J., Whittle, E., Lapchyk, L., Hildebrand, D., and Kachroo, P. (2007). The *Arabidopsis* stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. Plant Mol. Biol. *63*, 257-271.

24. Cahoon, E.B., Lindqvist, Y., Schneider, G., and Shanklin, J. (1997). Redesign of soluble fatty acid desaturase from plants for altered substrate specificity and double bond position. Proc. Natl. Acad. Sci. USA *94*, 4872-4877.

25. Joyard, J., Ferro, M., Masselon, C., Seigneurin-Berny, D., Salvi, D., Garin, J., and Rolland, N. (2010). Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. Prog. Lipid Res. *49*, 128-158.

26. Eugeni Piller, L., Besagni, C., Ksas, B., Rumeau, D., Bréhélin, C., Glauser, G., Kessler, F., and Havaux, M. (2011). Chloroplast lipid droplet type II NAD(P)H quinone oxidoreductase is essential for prenylquinone metabolism and vitamin K_1 accumulation. Proc. Natl. Acad. Sci. USA *108*, 14354-14359.

27. Shumskaya, M., Bradbury, L.M.T., Monaco, R.R., and Wurtzel, E.T. (2012). Plastid localization of the key carotenoid enzyme phytoene synthase is altered by isozyme, allelic variation, and activity. Plant Cell *24*, 3725-3741.

28. Lightner, J., Wu, J., and Browse, J. (1994). A mutant of *Arabidopsis* with increased levels of stearic acid. Plant Physiol. *106*, 1443-1451.

29. Sedeek, K.E.M., Qi, W., Schauer, M.A., Gupta, A.K., Poveda, L., Xu, S., Liu, Z.-J., Grossniklaus, U., Schiestl, F.P., and Schlüter, P.M. (2013). Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. PLoS One *8*, e64621.

30. Cahoon, E.B., Coughlan, S.J., and Shanklin, J. (1997). Characterization of a structurally and functionally diverged acyl-acyl carrier protein desaturase from milkweed seed. Plant Mol. Biol. *33*, 1105-1110.

31. Breitkopf, H., Onstein, R.E., Cafasso, D., Schlüter, P.M., and Cozzolino, S. (2015). Multiple shifts to different pollinators fuelled rapid diversification in sexually deceptive *Ophrys* orchids. New Phytol. *207*, 377-389.

32. Sedeek, K.E.M., Scopece, G., Staedler, Y.M., Schönenberger, J., Cozzolino, S., Schiestl, F.P., and Schlüter, P.M. (2014). Genic rather than genome-wide differences between sexually deceptive *Ophrys* orchids with different pollinators. Mol. Ecol. *23*, 6192-6205.

33. Xu, S., and Schlüter, P.M. (2015). Modeling the two-locus architecture of divergent pollinator adaptation: how variation in *SAD* paralogs affects fitness and evolutionary divergence in sexually deceptive orchids. Ecol. Evol. *5*, 493-502.

34. Wessinger, C.A., and Rausher, M.D. (2012). Lessons from flower colour evolution on targets of selection. J. Exp. Bot. *63*, 5741-5749.

35. Natarajan, C., Inoguchi, N., Weber, R.E., Fago, A., Moriyama, H., and Storz, J.F. (2013). Epistasis among adaptive mutations in deer mouse hemoglobin. Science *240*, 1324-1327.

36. Breen, M.S., Kemena, C., Vlasov, P.K., Notredame, C., and Kondrashov, F.A. (2012). Epistasis as the primary factor in molecular evolution. Nature *490*, 535-538.

37. Camps, M., Herman, A., Loh, E., and Loeb, L.A. (2007). Genetic constraints on protein evolution. Crit. Rev. Biochem. Mol. Biol. *42*, 313-326.

38. Harms, M.J., and Thornton, J.W. (2013). Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat. Rev. Genet. *14*, 559-571.

39. Ortlund, E.A., Bridgham, J.T., Redinbo, M.R., and Thornton, J.W. (2007). Crystal structure of an ancient protein: evolution by conformational epistasis. Science *317*, 1544-1548.

40. Liu, Q., Chai, J., Moche, M., Guy, J., Lindqvist, Y., and Shanklin, J. (2015). Half-of-the-sites reactivity of the castor Δ9-18:0-acyl carrier protein desaturase. Plant Physiol. *169*, 432-441.

41. Servedio, M.R., Van Doorn, G.S., Kopp, M., Frame, A.M., and Nosil, P. (2011). Magic traits in speciation: 'magic' but not rare? Trends Ecol. Evol. *26*, 389-397.

42. Xu, S., Schlüter, P.M., and Schiestl, F.P. (2012). Pollinator-driven speciation in sexually deceptive orchids. Int. J. Ecol. *2012*, Article ID 285081.

43. Nosil, P. (2012). Ecological speciation, (New York: Oxford University Press).

44. Sieber, P., Gheyselinck, J., Gross-Hardt, R., Laux, T., Grossniklaus, U., and Schneitz, K. (2004). Pattern formation during early ovule development in *Arabidopsis thaliana* Dev. Biol. *273*, 321-334.

45. Karlgren, A., Carlsson, J., Gyllenstrand, N., Lagercrantz, U., and Sundström, J.F. (2009). Non-radioactive *in situ* hybridization protocol applicable for Norway spruce and a range of plant species. J. Vis. Exp. *26*, 1205.

46. Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Prot. Sci. *8*, 978-984.

47. Sayle, R.A., and Milner-White, E.J. (1995). RASMOL: biomolecular graphics for all. Trends Biochem. Sci. *20*, 374-376.

48. Moche, M., Shanklin, J., Ghoshal, A., and Lindqvist, Y. (2003). Azide and acetate complexes plus two iron-depleted crystal structures of the di-iron enzyme ∆9 stearoyl-acyl carrier protein desaturase. J. Biol. Chem. *278*, 25072-25080.

49. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586-1591.

50. R Development Core Team (2012). R: A language and environment for statistical computing. Version 2.15.2. (Vienna, Austria: R Foundation for Statistical Computing).

51. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. Bioinformatics *20*, 289-290.

FIGURE LEGENDS

Figure 1. Desaturase Localization at the Transcript and Protein Levels

(A) Localization of *SAD2*, *SAD3* and *SAD5* in *Ophrys* flowers as detected by ISH. Zoom-in of *O. sphegodes* (*SAD2*) and *O. exaltata* (*SAD3*/*SAD5*) labellum sections (contrast adjusted), showing antisense and sense (control) probes (see also Fig. S1). The top cell layer represents the upper labellum surface. Strong signal is observed with antisense probes, especially in the epidermis, whereas only weak brownish background signal is observed in sense controls. (B) Subcellular localization of SAD5 protein based on transient expression of *Ophrys* SAD5-GFP in *N. benthamiana* protoplasts, visualized by confocal microscopy (other SADs in Fig. S2). From left to right, SAD5-GFP signal (green), chlorophyll autofluorescence (red), merged image (overlay) of SAD5-GFP and chlorophyll signal, and bright field image. Scale bar, 10 µm. The green SAD5-GFP signal shows a speckled pattern spatially associated with the chloroplasts.

Figure 2. GC/MS Analysis of SAD3 and SAD5 Desaturase Assays

(A-F) GC traces showing FA methyl esters (FAMEs) of assays with mass-labelled acyl-ACP substrates; unlabelled background contamination is derived from free FAs. FAME peak identities are explained in the insets. (A-C) Chromatograms showing assay of 7,7,8,8-D_4-16:0-ACP (D = ^{2}H, i.e. deuterium), (A) without desaturase, and with (B) SAD3, or (C) SAD5 desaturases. (D-F) Chromatograms showing assay of 18,18,18-D_3-18:0-ACP substrate (D) without desaturase, and with (E) SAD3, or (F) SAD5 desaturases. (G, H) MS fragmentation patterns of DMDS derivatives of FAMEs in (G) peaks 3 and (H) 6. This analysis confirms Δ^9 double-bond positions for both 16:1 and 18:1 desaturase reaction products. (I) Proposed major reaction catalysed by SAD5 and its link to 7-alkenes. (J) Corresponding scheme for C_{18} FAs/9-alkenes. (K) Percentage 16:0-ACP/18:0-ACP substrate converted (mean ± s.e.m.) by SAD3/SAD5 (other desaturases, Table S1).

Figure 3. Evolutionary Change in Desaturase Activity, Subcellular Localization, and Structure

(A) Phylogenetic relationships summarized after [15], with outgroup sequences from *Arabidopsis thaliana* (AtSSI2) and *Ricinus communis* (RcSAD). Enzymatic activities, subcellular localization and their inferred ancestral states are shown by symbols (see graphical inset legend): For enzyme activity on C_{16} and C_{18} FAs, double-bond insertion at positions Δ^9 or (for C_{16} FAs only) Δ^4 is indicated, or alternatively, a loss of catalytic function ('LOF'); '?' indicates uncertainty in ancestral character state or, in the case of C_{18} FAs, the effect of coding residual enzyme activity as either Δ^9 desaturation ('Δ9') or loss of function ('LOF'). For subcellular localization ('Loc'), 'cp' refers to localization associated with a diffuse (uniform), plastid-like signal and 'pg' with a speckled, plastoglobule-like signal, whereas '¿' indicates that inferred ancestral localization in the *SAD2*-clade depends on the subcellular localization of SAD1-A. (B) Homology models of protein structure. Amino acid side-chains within 6 Å of the modelled 18:0 substrate (molecule in large spheres) are shown in bold. Amongst these, for SAD2 and SAD5, amino acid differences (relative to SAD3) are highlighted in colour and inferred evolutionary changes at these sites indicated in the phylogeny. The two large dark yellow spheres represent the active site Fe atoms; smaller spheres are shown for atoms closest to (≤3 Å) 18:0 substrate.

Figure 4. Amino Acid and Activity Changes Along the Evolutionary Trajectory to SAD5-A

(A) Trajectory showing time points (t_{-3}-t_0); (B) configuration of amino acids in the substrate pocket and activities (as percentage substrate converted) of the corresponding resurrected proteins (Table S1; Fig. S3). (C) Pie charts indicating the predicted substrate conversion amounts (area) and ratios for homo- and heterozygotes (excluding LOF/LOF), assuming equal expression of both alleles and complete additivity in enzyme activities. The expected ratio effectively only changes after formation of a t_{-3}/t_0 heterozygote.

Figure 1

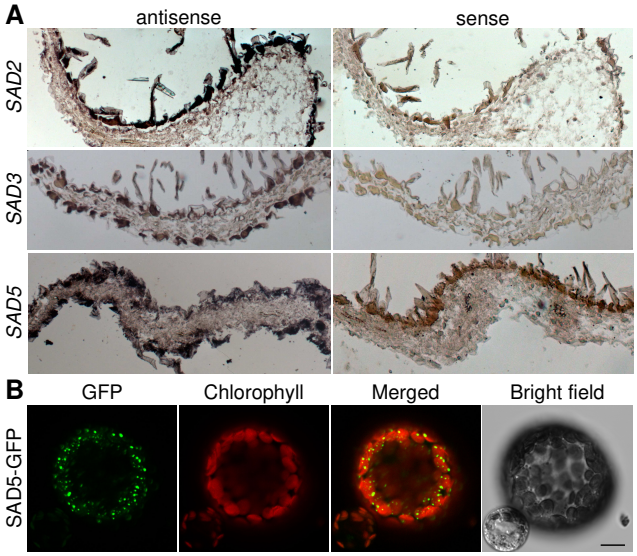

A
antisense
sense
SAD2
SAD3
SAD5
B
GFP
Chlorophyll
Merged
Bright field
SAD5-GFP

Figure 2

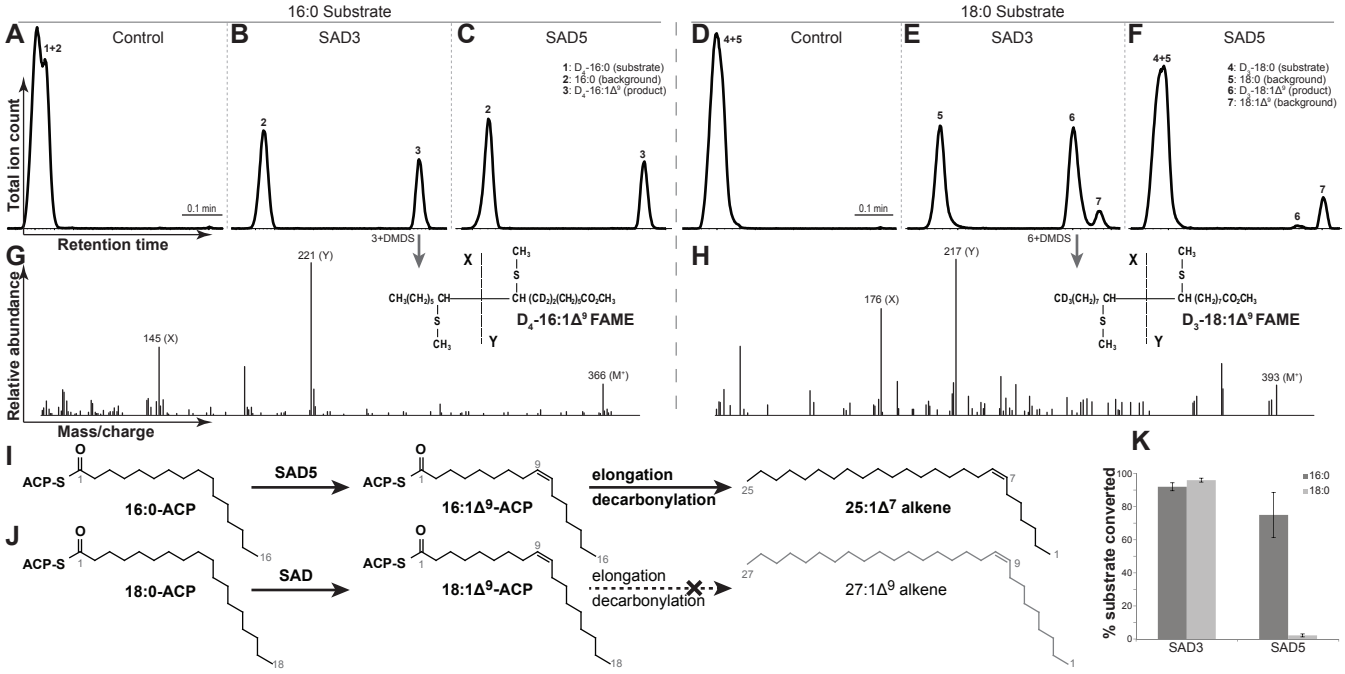

16:0 Substrate
18:0 Substrate
A Control
B SAD3
C SAD5
D Control
E SAD3
F SAD5
1: D4-16:0 (substrate)
2: 16:0 (background)
3: D4-16:1Δ9 (product)
4: D3-18:0 (substrate)
5: 18:0 (background)
6: D3-18:1Δ9 (product)
7: 18:1Δ9 (background)
Total ion count
Retention time
0.1 min
G
H
Relative abundance
Mass/charge
145 (X)
221 (Y)
366 (M+)
176 (X)
217 (Y)
393 (M+)
3+DMDS
6+DMDS
D4-16:1Δ9 FAME
D3-18:1Δ9 FAME
I
J
ACP-S
16:0-ACP
18:0-ACP
SAD5
SAD
16:1Δ9-ACP
18:1Δ9-ACP
elongation
decarbonylation
25:1Δ7 alkene
27:1Δ9 alkene
K
% substrate converted
16:0
18:0
SAD3
SAD5

Figure 3

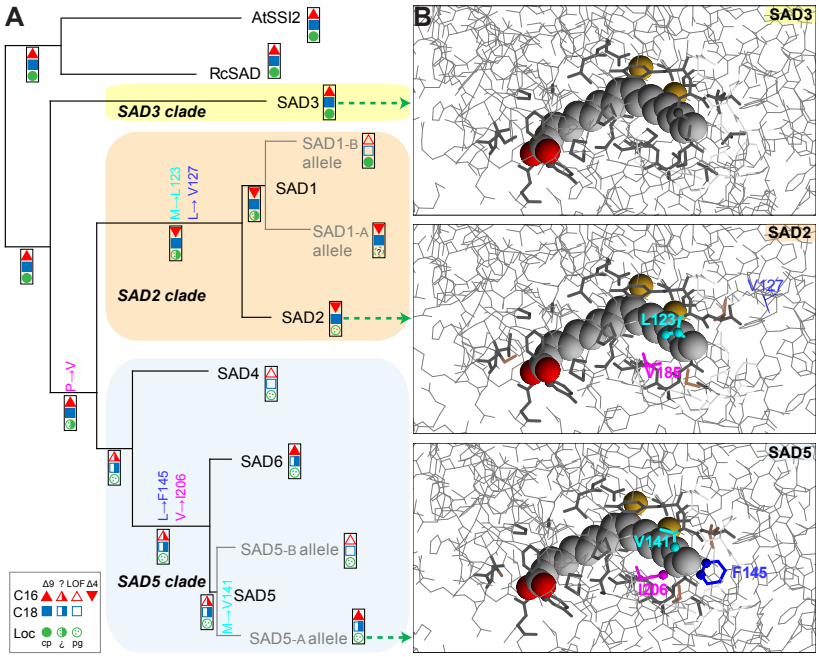
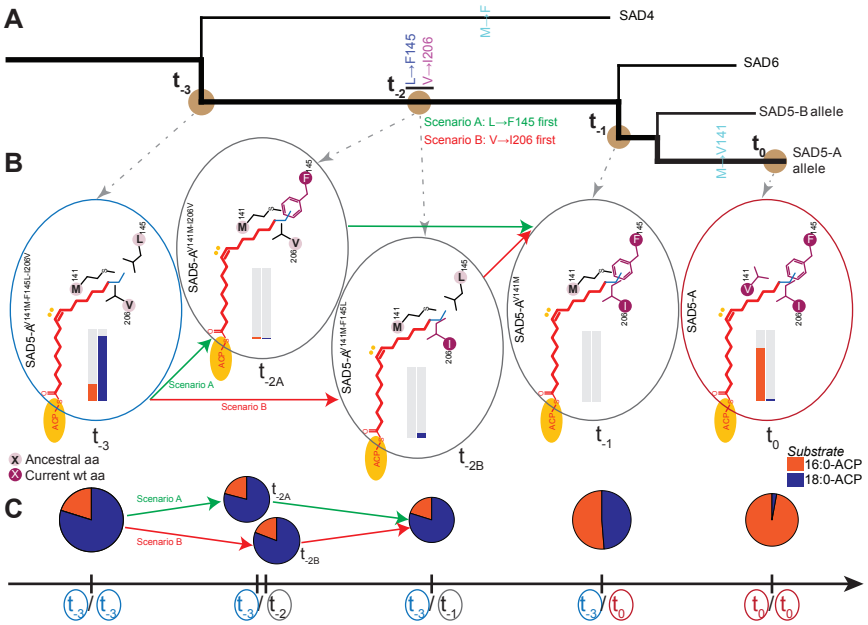

A
AtSSI2
RcSAD
SAD3
SAD3 clade
SAD1-B allele
SAD1
SAD1-A allele
SAD2
SAD2 clade
M→L123
L→V127
P→V
SAD4
SAD6
L→F145
V→I206
SAD5-B allele
SAD5
SAD5 clade
M→V141
SAD5-A allele
Δ9 ? LOF Δ4
C16
C18
Loc
cp ¿ pg
B
SAD3
SAD2
V127
L123
V185
SAD5
V141
F145
I206

Figure 4


A
SAD4
SAD6
SAD5-B allele
SAD5-A allele
M→F
L→F145
V→I206
M→V141
Scenario A: L→F145 first
Scenario B: V→I206 first
B
Scenario A
Scenario B
Ancestral aa
Current wt aa
Substrate
16:0-ACP
18:0-ACP
C
Scenario A
Scenario B

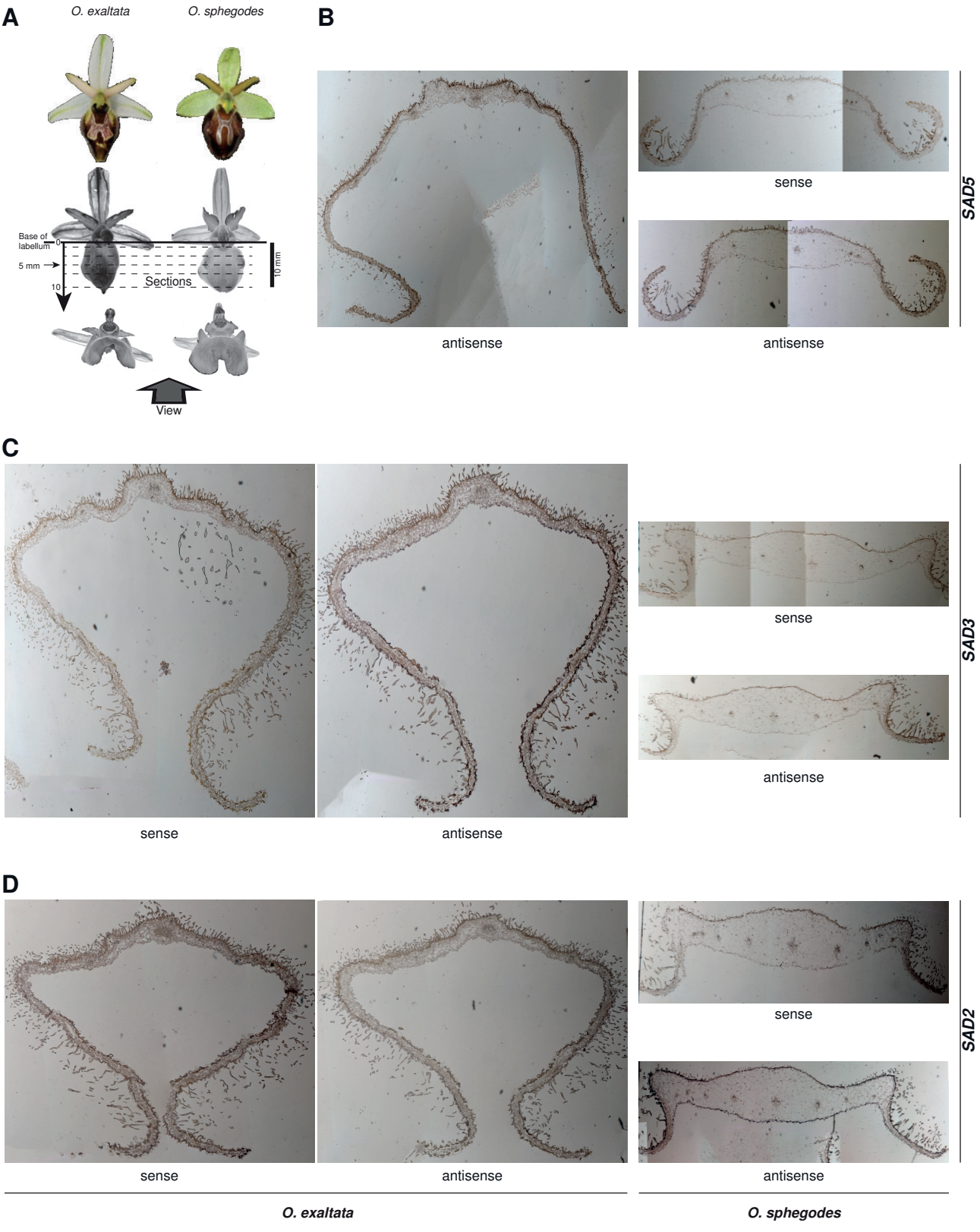

A
O. exaltata
O. sphegodes
Base of labellum
0
5 mm
10
Sections
10 mm
View
B
antisense
sense
antisense
SAD5
C
sense
antisense
sense
antisense
SAD3
D
sense
antisense
sense
antisense
SAD2
O. exaltata
O. sphegodes

Figure S1. Related to Figure 1.

RNA *in situ* hybridization of desaturases in *Ophrys* flowers. (A) Overview illustration showing *O. exaltata* and *O. sphegodes* flowers from top and front along with the layout and direction of view of thin sections (starting at the base of the labellum on top toward the apex at the bottom). (B-D) Thin sections of *O. exaltata* and *O. sphegodes* flowers at 5 mm from the top of the labellum, hybridized with sense (control) and antisense probes as indicated in the figure. It should be noted that qualitative absence of clear antisense signal (e.g. in non-epidermal layers) cannot be taken as evidence of absence of expression. (B) ISH of *SAD5*-clade desaturases (*SAD5* and *SAD6*) in *Ophrys* flowers**.** The antisense panel shows black signal in the epidermal layer of *O. exaltata*, but not in *O. sphegodes. O. exaltata* antisense signal is stronger than the brownish background seen with the sense and antisense probes in *O. sphegodes*. This is consistent with the expression pattern reported for *SAD5* [S1], which was highly expressed in the labellum of *O. exaltata* but not in *O. sphegodes*. Since *SAD5* was more highly expressed than *SAD6* in *O. exaltata* [S1], ISH signal is expected to be primarily due to *SAD5*. (C) ISH of *SAD3* (the only *SAD3*-clade member) in *Ophrys* flowers. Antisense signal was detected in the epidermal layer of both species as compared to the sense control, although this is much clearer in *O. exaltata*, where a strong antisense signal is also detected in the abaxial (inner) epidermis of the labellum. Previous gene expression analyses showed *SAD3* to be ubiquitously expressed in different plant tissues and developmental stages, suggesting it to encode a housekeeping desaturase [S1, 3]; these expression data are consistent with ISH despite low signal intensity in *O. sphegodes*. (D) ISH of *SAD2*-clade desaturases (*SAD1* and *SAD2*) in *Ophrys* flowers. The antisense panel shows blackish signal in the epidermal layer of *O. sphegodes* that is clearly stronger than the brownish background signal observed in the sense control. However, no such difference was observed between sense and antisense signals in *O. exaltata* epidermis. The difference between *O. sphegodes* and *O. exaltata* is consistent with previous gene expression analyses [S1, 3] that showed *SAD2*-clade genes to be highly expressed in mature *O. sphegodes* labella, whereas expression in *O. exaltata* labella was low. The higher level of *SAD2* expression as compared to *SAD1* expression in *O. sphegodes* [S1, 3] suggests that ISH signal in *O. sphegodes* is primarily due to *SAD2*. Full high-resolution ISH data for *SAD2*, *SAD3* and *SAD5* probes from both species (including thin sections at 1, 3, 5, 7 and 10 mm from the base of the labellum) are available from the Dryad data archive, doi: 10.5061/dryad.698s5.

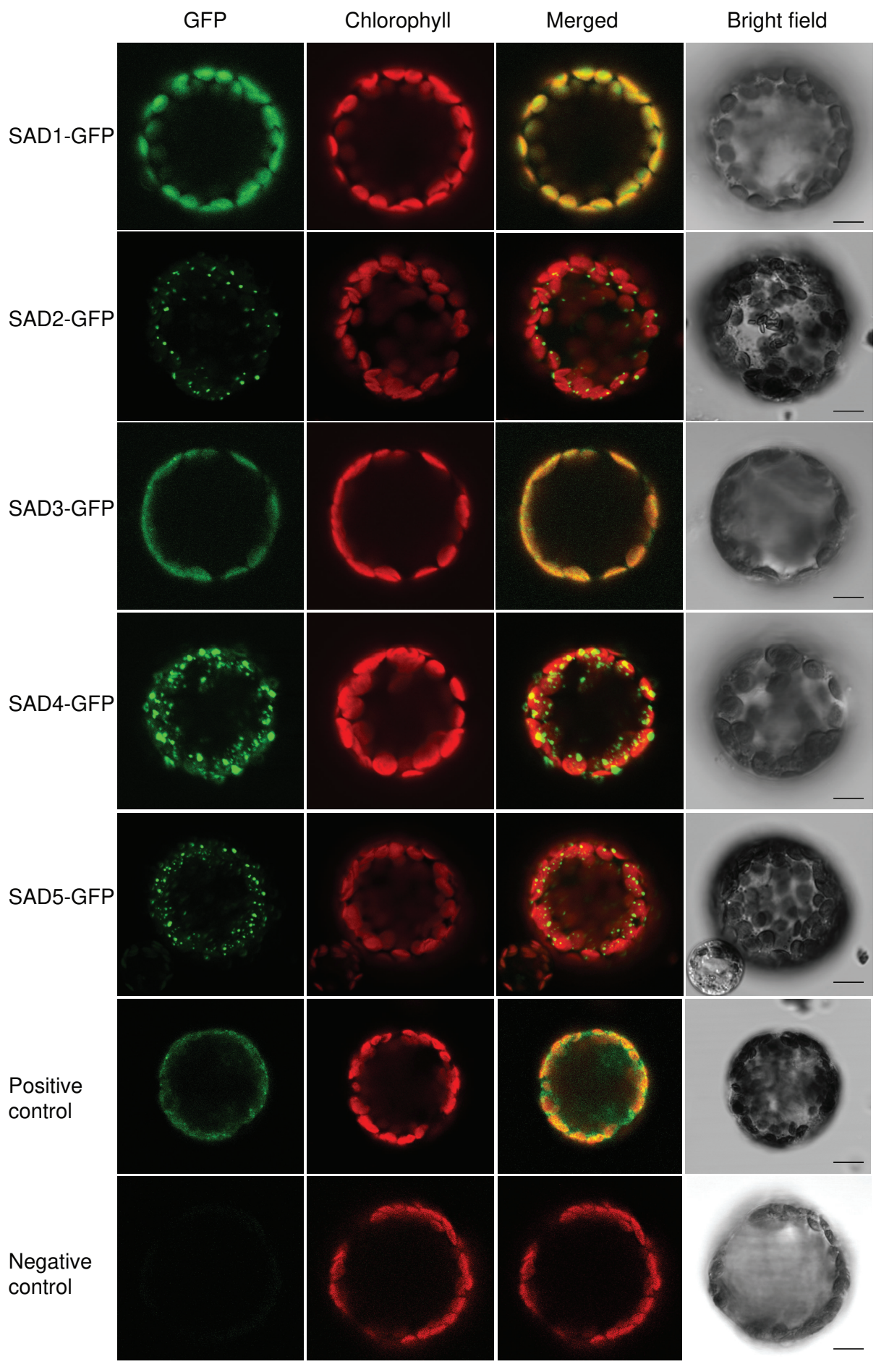

GFP
Chlorophyll
Merged
Bright field
SAD1-GFP
SAD2-GFP
SAD3-GFP
SAD4-GFP
SAD5-GFP
Positive control
Negative control

Figure S2. Related to Figure 1.

Subcellular localization of SAD proteins. Subcellular localization was assessed by transient expression of *Ophrys* SAD-GFP fusion constructs in *Nicotiana benthamiana* leaf protoplasts. Fluorescence is shown as imaged by laser scanning confocal microscopy. The four columns show (from left to right): GFP signal (green channel), chlorophyll autofluorescence (red channel), merged GFP and chlorophyll signals (overlay of both panels), and bright field images. PPH-GFP [S5] was used as a positive control for showing signal in the chloroplast, and untransformed *Agrobacterium* as a negative control. The green GFP signals of the orchid *SAD* constructs were associated with plastids. For *SAD1* and *SAD3* constructs, signal distribution was smooth and similar to the positive control, whereas signal of *SAD2*, *SAD4*, and *SAD5* was speckled, similar to speckles reported for plastoglobule localization [S6, 7]. SAD-GFP subcellular localization experiments were repeated four times per construct, except for *SAD1* and *SAD3* (three replicates). The scale bars shown represent 10 µm.

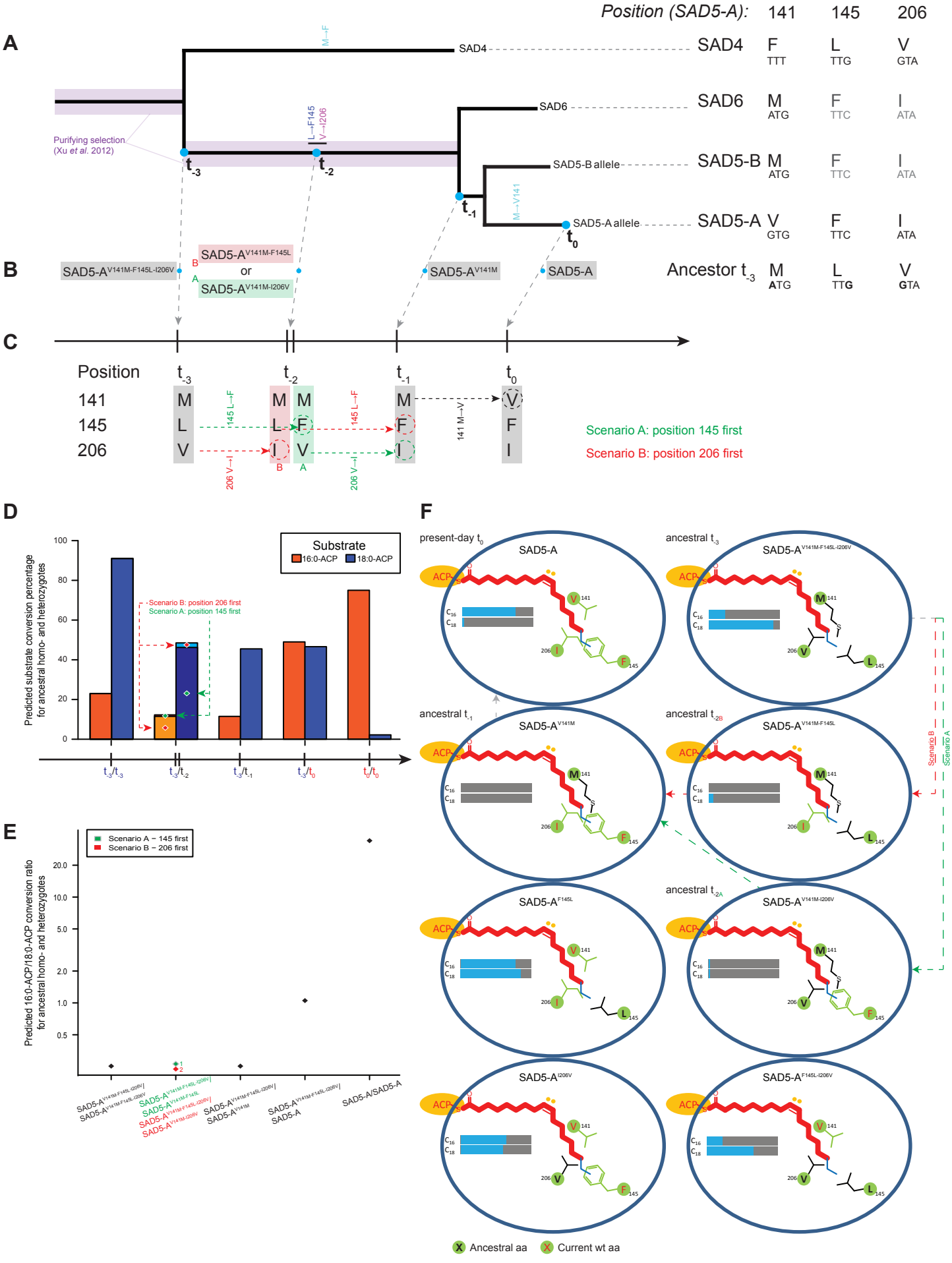

Position (SAD5-A): 141 145 206
A
SAD4 F TTT L TTG V GTA
SAD6 M ATG F TTC I ATA
SAD5-B M ATG F TTC I ATA
SAD5-A V GTG F TTC I ATA
Ancestor t-3 M ATG L TTG V GTA
M→F
L→F145
V→I206
M→V141
Purifying selection (Xu et al. 2012)
SAD4
SAD6
SAD5-B allele
SAD5-A allele
t-3
t-2
t-1
t0
B
SAD5-AV141M-F145L-I206V
SAD5-AV141M-F145L
or
SAD5-AV141M-I206V
SAD5-AV141M
SAD5-A
C
Position
141
145
206
t-3 M L V
t-2 M L I / M F V
t-1 M F I
t0 V F I
145 L→F
206 V→I
141 M→V
Scenario A: position 145 first
Scenario B: position 206 first
D
Predicted substrate conversion percentage for ancestral homo- and heterozygotes
Substrate
16:0-ACP
18:0-ACP
Scenario B: position 206 first
Scenario A: position 145 first
t-3/t-3
t-3/t-2
t-3/t-1
t-3/t0
t0/t0
E
Predicted 16:0-ACP/18:0-ACP conversion ratio for ancestral homo- and heterozygotes
Scenario A – 145 first
Scenario B – 206 first
SAD5-AV141M-F145L-I206V
SAD5-AV141M-F145L
SAD5-AV141M-I206V
SAD5-AV141M
SAD5-A
SAD5-A/SAD5-A
F
present-day t0
SAD5-A
ancestral t-3
SAD5-AV141M-F145L-I206V
ancestral t-1
SAD5-AV141M
ancestral t-2B
SAD5-AV141M-F145L
ancestral t-2A
SAD5-AF145L
SAD5-AV141M-I206V
SAD5-AI206V
SAD5-AF145L-I206V
ACP
Scenario B
Scenario A
X Ancestral aa
X Current wt aa

Figure S3. Related to Figure 4.

Schematic diagram of substrate and amino acid configuration in the SAD5-A variants analysed in an evolutionary context. Wild-type, single, double and triple amino acid replacement SAD5-A protein variants were analysed. Several of these protein variants represent amino acid configurations around the substrate binding pocket of ancestral proteins along the evolutionary trajectory leading to present-day SAD5-A. (A) Phylogenetic relationships in the *SAD5* clade after ref. [S1], showing amino acids and codons in the different SAD proteins at residues homologous to positions 141, 145 and 206 of SAD5-A. Inferred amino acid change at these positions is indicated in the phylogeny; branches with purifying selection significantly stronger than the background rate [S1] are also highlighted. Along the branches leading to SAD5-A, four blue circles indicate three time-points in the evolutionary history of SAD5 (corresponding to ancestral SAD5-like proteins t_{-3}, t_{-2} and t_{-1}) and present-day wild-type protein variant SAD5-A (t_0). (B) Identity of the corresponding protein variants assayed *in vitro* (see Table S1). The exact order of mutations at positions 145 and 206, and hence the corresponding amino acid configuration of ancestral protein t_{-2}, are unknown, giving rise to two evolutionary scenarios of change. (C) Scenario A (green) assumes that change at position 145 (change from L145 to F145) occurred first, whereas scenario B (red) assumes that change at position 206 (V206 to I206) occurred first. (D) A possible route of evolutionary change and the associated hypothesized net desaturase activities (based on values in Table S1) in homozygous and heterozygous plants assuming an equal expression of both alleles and complete additivity in enzyme activities (LOF/LOF combinations not shown). Blue and red bars indicate the predicted percentage of substrate conversion for C_{16} and C_{18}, respectively. Both evolutionary scenarios are shown and have very similar predictions of net enzyme activity. (E) Change of expected 16:0-ACP/18:0-ACP conversion ratio (*log*-scale) over evolutionary time, based upon panel (D), again showing both scenarios of change at position 145 and 206 with very similar ratios. The expected ratio effectively only changes after a t_{-3}/t_0 heterozygote is formed. (F) Amino acid configuration and percentage C_{16} and C_{18} substrate converted in the SAD5-A variants analysed *in vitro*: present-day wild-type SAD5-A (t_0), SAD5-A$^{V141M-F145L-I206V}$ (ancestral configuration t_{-3}), SAD5-A^{V141M} (ancestral configuration t_{-1}), SAD5-A$^{V141M-F145L}$ (ancestral configuration t_{-2B}, assuming scenario B), SAD5-A^{F145L}, SAD5-A$^{V141M-I206V}$ (ancestral configuration t_{-2A}, assuming scenario A), SAD5-A^{I206V} and SAD5-A$^{F145L-I206V}$. The first 16 carbons of a substrate fatty acyl group are shown in bold red line, whereas the remaining two carbons (carbon 17 and 18) are shown in a thin blue line. Three amino acid side chains at positions 141, 145 and 206 of SAD5-A are depicted in green if identical to wild-type SAD5-A (associated with red letters) or black (associated with bold, black letters) if replaced with the inferred ancestral amino acid at this position. Blue bars indicate the mean percentage of conversion of C_{16} and C_{18} FA substrates measured for the respective protein variants in desaturase activity assays (Table S1).

Supplemental Tables

Table S1. Related to Figures 2 and 4.

Enzymatic activities of orchid SAD proteins and mutants. Enzyme activity is shown as the percentage of conversion of saturated to unsaturated fatty acid intermediates in GC/MS-based assays, and the ratio of the activities towards the two substrates is shown as a measure of catalytic specificity. Data shown are from 3 replicates (except where indicated otherwise) and presented as mean ± standard error, except where specific reaction product (see Fig. 2) was not detected (n.d.). SAD5-A refers to one of two (allelic) isoforms of SAD5, the other isoform likely being non-functional [S1].

Enzyme	**Conversion percentage [%]**		**Ratio**
Substrate:	*16:0-ACP*	*18:0-ACP*	C_{16}/C_{18}
Paralogues			
SAD3	92.0 (± 2.4)	96.0 (± 1.2)	0.96
SAD4	n.d.	n.d.	—
SAD5-A	75.0 (± 14.0)[a]	2.2 (± 0.9)	**34.09**
Mutants			
SAD5-A^{V141M}	n.d.	n.d.	—
SAD5-A^{F145L}	77.0 (± 3.4)	85.0 (± 4.8)	0.91
SAD5-A^{I206V}	64.0 (± 3.1)	59.0 (± 1.5)	1.08
SAD5-A$^{V141M-F145L}$	n.d.	6.0 (± 1.5)[a]	—
SAD5-A$^{V141M-I206V}$	1.5 (± 0.1)	1.4 (± 0.7)[a]	1.07
SAD5-A$^{F145L-I206V}$	22.0 (± 5.2)	66.0 (± 5.5)[a]	0.33
SAD5-A$^{V141M-F145L-I206V}$	23.0 (± 5.1)	91.0 (± 4.2)	**0.25**

[a] From two replicates.

Table S2. Related to Experimental Procedures.

Predicted protein characteristics of selected orchid desaturases. Protein length is given in amino acids, including initiation methionine and putative plastid transit peptide (TP). The predicted presence of a TP is indicated by a 'Y' in the TP column, with the ChloroP prediction score in round brackets, and the predicted length listed in the respective column. Average molecular weight (MW) and isoelectric point (pI) as predicted by the ExPASy server [S2] are shown for the proteins with (+) and without (-) the TP. Similar data for SAD1-B, SAD2 and SAD3 can be found in Table S5 of ref. [S3].

Protein/allele group	**Accession**	**Length**	**TP (score)**	**TP length**	**MW +TP**	**MW -TP**	**pI +TP**	**pI -TP**
SAD1-A	JN413063	375	Y (0.503)	40	42554.5	38206.5	6.28	5.71
SAD4-A	JN412905	388	Y (0.554)	30	44464.8	41267.0	7.63	6.64
SAD4-A	KM485309	387	Y (0.549)	30	44364.7	41166.9	7.18	6.47
SAD4-A_2	KM485308	388	Y (0.559)	30	44359.6	41157.8	7.18	6.47
SAD5-A	JX307700	393	Y (0.549)	32	45167.3	41661.4	6.91	6.47
SAD5-B	JX307701	389	Y (0.545)	32	44781.0	41366.2	6.9	6.47
SAD5-B	KM485307	389	Y (0.545)	32	44875.1	41460.3	7.7	6.87
SAD6-B	JX307699	393	Y (0.544)	32	45207.5	41810.6	6.91	6.47

Table S3. Related to Figure 3.

Putatively ACP-interacting amino acids based on protein structure. This table lists amino acids of mature *Ricinus communis* desaturase (mRcSAD) that interact with stearoyl-ACP in a crystal structure [S4], along with the corresponding amino acids in the *Ophrys* desaturases and their positions in unprocessed SAD sequence (same accessions as used for homology modelling, see Table under Supplemental Experimental Procedures: Sequence selection). For amino acids that were polymorphic in the set of sequences analysed (see the same Table), variants are listed in descending order of occurrence.

mRcSAD	**Amino acid in *Ophrys* SAD**						**Position in *Ophrys* SAD**		
	SAD1	SAD2	SAD3	SAD4	SAD5	SAD6	SAD2-A	SAD3	SAD5-A
R260	R/Q[a]	R/Q[a]	K	R	R	R	266	294	287
K262	R	H/R[b]	K	K	K	K	268	296	289
D280	T	T	K	N	D	D	286	314	307
R333	V	V	R	R	R	R	340	367	361
R336	K	K	K	K	K	K	343	370	364
L337	M/V/L[c]	M/L[c]	V	L	L	L	344	371	365
R340	R	R	R	R	R	R	347	374	368
R344	R	R/H[d]	M	R	R	R	351	378	372

[a] Polymorphic R/Q in SAD1-A, Q in SAD1-B; R in SAD2, except for Q in SAD2-C.
[b] H in SAD2-A.
[c] Polymorphic V/M in SAD1-A, L in SAD1-B; M in SAD2, except for L in SAD2-B.
[d] H only in SAD2-B.

Supplemental Experimental Procedures

Sequence selection

A representative selection of naturally occurring sequence diversity at the loci *SAD1–SAD6*, from data previously obtained by Xu et al. [S1], was compiled for evolutionary and functional analysis. This sequence selection covered all known *SAD* allele groups and observed *SAD* nucleotide variation, but care was taken to exclude sequences that were only observed once in order to minimize the chance of including laboratory-generated variation (e.g. due to PCR or sequencing errors) in the analysis. The resulting set of sequences (listed in the Table below) contained 59 *Ophrys* desaturase sequences for evolutionary analysis and three well-characterized non-orchid reference genes. Further molecular and computational analyses were carried out on common and putatively functional alleles [S1] with a representative sequence composition that were part of the set of selected sequences (plus the addition of two additional *SAD5*/*SAD6* sequences, see Table below). The respective source sequences for molecular analyses had previously been cloned into pDRIVE (Qiagen) vector [S1].

Table listing sequences selected for evolutionary analysis. Allele group designations follow ref. [S1], all 'A' allele groups being putatively functional. Many alleles are shared across several *Ophrys* species [S1]; hence, species of sequence origin is not indicated here. Sequences used for additional purposes are marked with italicized superscripts enclosed in angle brackets: $^{<B>}$, biochemical analysis; $^{<G>}$, GFP fusion; $^{<P>}$, bioinformatic prediction of transit peptide and subcellular localization; $^{<M>}$, homology modelling; $^{<R>}$, RNA *in situ* hybridization; lower case letters ($^{<b>}$, $^{<p>}$, $^{<m>}$) denote previously published analyses [S3]. In some cases where GenBank accessions were nearly identical, a consensus sequence of several GenBank accessions was used (indicated by the '+' sign between accessions).

Locus (N)	Allele group (N)	Sequences used (GenBank accessions)
Outgroup (3)		*Arabidopsis thaliana SSI2* (NM_180069)[a], *Hedera helix* Δ^4 acyl-ACP desaturase (AY998047), *Ricinus communis SAD* (M59857)
***SAD1* (8)**	A (3)	JN413063$^{<B,P,M>}$, JN413064, JN413078
	B (5)	FR688108$^{<b,G,p,m>}$, JN413036, JN413043, JN413045, JN413052
***SAD2* (9)**	A (4)	FR688109$^{<b,G,R,p,m>}$, JN412989, JN412991, JN412992
	B (2)	JN413007, JN413008
	C (1)	FR688110$^{<b,p,m>}$
	D (1)	JN412988[b]
	E (1)	JN412990
***SAD3* (16)**	A (16)	FR688106$^{<p,m>}$, FR688107, JN412955, JN412956, JN412957$^{<B,G,R>}$,[c], JN412958, JN412959, JN412960, JN412961[d], JN412970+JN412971+JN412972, JN412974, JN412976, JN412977, JN412984, JN412985, JN412986
***SAD4* (14)**	A (9)	JN412899, JN412900, JN412903, JN412905$^{<P,M>}$, JN412907$^{<B,G>}$, JN412908, JN412909, JN412910, KM485309$^{<P,M>}$
	A_2 (5)	JN412901, JN412902, JN412904, JN412906, KM485308[P,M]
***SAD5* (7)**	A (3)	JN412888, JN412893$^{<B,R>}$, JX307700$^{<P,M,G>}$,[e]
	B (4)	JN412882+JN412883+JN412884, JX307701$^{<P,M>}$, KM485307$^{<P,M>}$, KM485306
***SAD6* (5)**	A (1)	JN412880$^{<B,M>}$
	B (4)	JN412858, JN412859 (JN412861$^{<B>}$)[f], JN412869, JX307699$^{<P,M>}$,[e]

[a] *Arabidopsis* gene At2g43710.
[b] This sequence contains a long insertion, which was deleted from the alignment that was used for analysis.
[c] This is the most similar sequence to FR688106, which was not available as a clone.
[d] This accession contains a stretch of several ambiguous nucleotides (N), which was replaced by the *SAD3* consensus sequence for the purpose of analysis.
[e] The actual sequence used for the GFP fusion is accession KP184494 (cloned from cDNA; see Supplemental Experimental Procedures), which matches the N-terminal amino acid sequence of SAD5-A JX307700 (including the putative transit peptide) perfectly, except for two amino acids at positions 8 and 9 (KP184494: FS; JX307700: ST); the C-terminal amino acid sequence of KP184494 is very similar to SAD6-B JX307699.
[f] The clone in round brackets differs by a single (unambiguous) nucleotide and was used for making the expression clone.

Molecular cloning

Table listing the oligonucleotides used in this study.

Primer name	Oligonucleotide sequence (5′→3′)	Purpose
SAD2-ISH1f	CCGCCGACCACCTCCTCC	Preparation of pDRIVE-SAD constructs, PCR screening and Sanger sequencing.
SAD2-ISH1r	CTGTTGCCTACTCCAACGTCC	
SAD3-ISH1f	TCGGGCTATGCCGATTGTAA	
SAD3-ISH1r	ATCTGTCTCATGTCTACTCTTC	
SAD5-ISH1f	CATAAGATGGAGATCTTCAAGTC	
SAD5-ISH1r	CTGGAAGGAGGTGTAGATGAAT	
M13f (-20)[a]	GTAAAACGACGGCCAGT	
BN~mSAD1A_f	AAAAACGTCTCACATGAGAAACTTCGCAACGAACAGCA	Preparation of pET9d-mSAD constructs, PCR-screening and Sanger-sequencing.
SAD1A~H_r	GTGATGGTGGTGATGCACGCGGACCTGCTTGTTGAA	
BN~mSAD2A_f	AAAAACGTCTCACATGGCAGCCATCAACAGCACTCAT	
SAD2A~H_r	GTGATGGTGGTGATGCACGCAGACTTGCTTATTGAA	
BN~mSAD3_f	AAAAACGTCTCACATGGCTATGCCGATTGTAATGAAA	
SAD3~H_r	GTGATGGTGGTGATGAAGCATCACTTCCCGGTCATG	
BN~mSAD4_f	AAAAACGTCTCACATGTCCTCCATGTCAGTCCACCAT	
SAD4~H_r	GTGATGGTGGTGATGAACCTGTACTTCCCTATCAAA	
BN~mSAD5A_f	AAAAACGTCTCACATGTCCTCCATGAACGCAAGCCGG	
SAD5A+6~H_r	GTGATGGTGGTGATGAACCTGCACTTCCCTGTTGAA	
BBH-Ad_r	TTTTTCGTCTCTGATCTTAGTGGTGATGGTGGTGATG	
pDONR207f	ATGCCTGGCAGTTCCCTACTCTCG	Preparation of pMDC83-SAD-GFP constructs, PCR-screening and Sanger-sequencing.
pDONR207r	AATGTAACATCAGAGATTTTGAGACACG	
ptB1fs-SAD1f (M)[b]	AAAAAGCAGGCTTTATGGAACTTCACCTTGCACTACT	
ptB2fs-SSAD1r	AGAAAGCTGGGTACACGCGGACCTGCTTGTTGAATATCC	
ptB1fs-SAD2f (M)[b]	AAAAAGCAGGCTTTATGGAACTTCACCTTGCACTACG	
ptB2fs-SSAD2r	AGAAAGCTGGGTACACGCAGACTTGCTTATTGAATATCC	
pB1-SAD3_5' (*M)	AAAAAGCAGGCTAAATGGCTTTGAGATCGCTTTTCC	
fpB2-SAD3_3'-1r	AGAAAGCTGGGTAAAGCATCACTTCCCGGTCATG	
pB1-SAD4_5'-1f*M	AAAAAGCAGGCTAAATGGCAATGATCGGCTTTCCAACA	
fpB2-SAD4_3'-1r	AGAAAGCTGGGTAAACCTGTACTTCCCTATCAAA	
pB1-SAD5+6_5'_1f (*M)	AAAAAGCAGGCTAAATGGCTTTCAGCAGCTCTTCTTCCAC	
fpB2-SAD5+6_3'_1r	AGAAAGCTGGGTAAACCTGCACTTCCCTGTTGAAAACCCA	
AttB1AdP[c]	GGGGACAAGTTTGTACAAAAAAGCAGGCT	
AttB2AdP[c]	GGGGACCACTTTGTACAAGAAAGCTGGGT	

[a] As published previously [S8].
[b] As published previously [S3].
[c] Sequence as specified by Invitrogen.

pDRIVE-SAD constructs for ISH

Three pDRIVE constructs containing *SAD2*, *SAD3* and *SAD5* were used as representatives of the three main *Ophrys* SAD lineages [S1], as listed in Supplemental Experimental Procedures (Table detailing sequence selection). The maximally divergent region between the three *SAD* clades was amplified by PCR from these clones (*SAD2*: 516 bp, *SAD3*: 513 bp, *SAD5*: 486 bp; for primer sequences see Table above) using DreamTaq Green PCR Master Mix (Thermo Scientific); the PCR conditions were: 94°C 2 min, 35× (94°C 30 s, 50°C 30 s, 72°C 30 s), 72°C 5 min. The PCR products were analysed on 1.2% agarose gel and bands at the expected product size (486-516 bp) were purified from the gel using the NucleoSpin Gel and PCR Clean-up kit (Macherey Nagel). The cleaned fragments were cloned into pDRIVE vector using QIAGEN PCR Cloning Kit and the manufacturer's protocol and transformed into *Escherichia coli* TOP10 chemically competent cells (Life Technologies). Cells were grown overnight on Luria-Bertani (LB) agar plates containing 100 µg/ml ampicillin, 50 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 80 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). White colonies were screened by PCR using combinations of primers (M13f/SAD-ISH1f and M13f/SAD-ISH1r; see Table above) and Sanger sequenced (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems) to verify sequence and select clones in sense and antisense directions for ISH probe preparation.

pMDC83-SAD-GFP constructs for subcellular localization

Full-length sequences of *SAD1–SAD4* (~1100 bp) were re-amplified by PCR from clones (see Supplemental Experimental Procedures: Tables listing selected sequences and oligonucleotides) using Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific) to produce fragments for Gateway-based cloning (Invitrogen). PCR conditions were: 98°C 30 s, 10× (98°C 10 s, 55°C 30 s, 72°C 45 s), 72°C 5 min followed by another 25 cycles with *att*B1 and *att*B2 adapter primers (AttB1AdP/AttB2AdP, see Table listing oligonucleotides above). Since no full-length clones were available for *SAD5*/*SAD6* (all available clones were incomplete at the 5' end), a full-length sequence (GenBank accession KP184494) was amplified from *O. exaltata* cDNA [S1] (accession PMS403D from the Foce Garigliano population) to make a SAD5-GFP fusion construct, using a primer (Table listing oligonucleotides above) binding at the start of the predicted coding sequence as previously determined in a 5' RACE experiment [S1]. No separate SAD6-GFP construct was made because of SAD6's high similarity to SAD5, especially at the predicted N-termini of the unprocessed proteins: (i) the first >50 amino acids translated from cloned partial *SAD5-A* and *SAD6* sequences (Supplemental Experimental Procedures: Table listing selected sequences) are identical; (ii) the 5' RACE data previously obtained for *SAD6* are all from seemingly non-functionalized *SAD6-B* alleles [S1] and the sequences moreover almost identical to *SAD5* 5' ends; (iii) also, cDNA cloning did not yield additional cloned sequences. For all SAD-GFP fusions, PCR products were analysed on 1.2% agarose gel and purified from the gel using NucleoSpin Gel and PCR Clean-up kit (Macherey Nagel). The cleaned *att*B-site-containing PCR fragments were recombined with pDONR207 vector (Invitrogen) using BP clonase II (Invitrogen), which generated pENTR clones, according to the manufacturer's protocol. The pENTR clones were transformed into *E. coli* TOP10 competent cells and grown on LB agar with gentamicin (10 µg/ml). These clones were PCR-screened with vector-located primers (Table listing oligonucleotides above), miniprepped (Thermo Scientific GeneJET Plasmid Miniprep Kit) and Sanger sequenced. An LR reaction using LR clonase II (Invitrogen) was performed with the resulting plasmids and a GFP-containing binary expression vector (pMDC83) [S9] under the control of *Cauliflower mosaic virus (CaMV) 35S* RNA promoter. This allowed C-terminal fusion of full-length SADs with GFP. The constructs (pMDC83-SAD-GFP) were transformed into *E. coli* TOP10 and grown on LB agar with kanamycin (50 µg/ml), then PCR-screened using *SAD*-specific primers (Table listing oligonucleotides above), miniprepped (Thermo Scientific GeneJET Plasmid Miniprep Kit) and Sanger sequenced. After sequence verification, the pMDC83-SAD-GFP constructs were transformed into *Agrobacterium tumefaciens* (strain GV3101) and the bacteria grown on LB agar with kanamycin (50 µg/ml), rifampicin (25 µg/ml) and gentamicin (25 µg/ml). Colonies carrying the construct were selected by PCR screening and sequences again verified by Sanger sequencing.

pET9d-mSAD constructs for biochemical analysis

Constructs for expression analysis of mature SAD (mSAD) protein for biochemical characterization were initially engineered to contain an N-terminal His_6-tag followed by a tobacco etch virus (TEV) protease cleavage site and cloned into pRSET-GW [S3] vector, but since tag removal by TEV protease cleavage failed and proteins were deemed to be non-functional as a result, an alternative set of expression clones was prepared as follows.

The mature sequence (full length without the putative transit peptide) of *SAD1-A* and *SAD3–SAD6* (~1100 bp) was re-amplified from clones (Supplemental Experimental Procedures: Tables listing selected sequences and oligonucleotides). Six C-terminal His residues (followed by a stop codon) were added by PCR in place of the orchid *SADs'* stop codons. In order to facilitate restriction enzyme (RE)-based cloning into pET9d despite internal RE sites in the *SAD* PCR products, *Nco*I- and *Bam*HI-compatible ends were generated during PCR amplification: The primers used for amplification contained recognition sites for *Bsm*BI (which cuts outside of its recognition sequence), so that after digestion, *Nco*I- and *Bam*HI-compatible ends were left on the PCR product. PCR was carried out using Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific) in 50 µl reactions containing ~5-10 ng source plasmid, 1 u Phusion Polymerase, 1× Phusion buffer, 1 µl 10 mM each dNTP, 0.5 µl 5 µM *SAD* reverse primer (with partial His-tag), 2 µl 5 µM *SAD* forward primer and 2 µl 5 µM BBH-Ad_r primer (see Table listing oligonucleotides above). Cycling conditions were: 98°C 30 s, 30× (98°C 5 s, 60°C 15 s, 72°C 30 s), 72°C 3 min. The PCR products were cleaned as described above and digested using FastDigest *Bsm*BI (Thermo Scientific) for 8 min at 37°C, followed by inactivation of the enzyme. The digested products were ligated with *Nco*I/*Bam*HI-digested pET9d vector (Novagen) using T4 DNA ligase (Thermo Scientific) at 17°C overnight. The resulting constructs (pET9d-mSAD) were chemically transformed into *E. coli* BL21-Gold (DE3) competent cells (Agilent Technologies) to obtain a high level of recombinant protein expression. Colony PCR and Sanger sequencing were carried out to select clones without any mutations.

To investigate the functional importance of three amino acids in mature SAD5-A (sequence of accession JN412893) in the context of the protein's evolution, variant recombinant proteins were engineered in which residues of interest were replaced by their predicted ancestral amino acids (see Fig. S3). To do so, point

mutations were introduced relative to pET9d-mSAD5A by a commercial mutagenesis service (GenScript, Piscataway, NJ, USA): (i) V141M, reverting SAD5-A's V141 (GTG codon) into the ancestral M (ATG codon); (ii) F145L, reverting SAD5-A's F145 (codon TTC) into the ancestral L (codon TTG); (iii) I206V, reverting SAD5-A's I206 (codon ATA) into the ancestral V (codon GTA); SAD5 amino acid residue numbering is based on the (unprocessed) translated full-length sequence of *SAD5-A* accession JX307700. In addition to the single amino acid replacements, corresponding constructs for the triple replacement and all double amino acid replacement combinations were obtained from GenScript.

RNA in situ *hybridization*

In situ hybridization was performed to locate the expression of *SAD*s in flower labellum tissue. For the three *SAD* clades [S1], the most divergent clade-specific ~500 bp of sequence of *SAD2, SAD3* and *SAD5* were subcloned into pDRIVE (QIAGEN PCR Cloning Kit) in sense and antisense direction, as detailed in earlier in Supplemental Experimental Procedures. Each construct (5 µg) was linearized using 100 u of *Hin*dIII (New England Biolabs) for 4 h. RNA probes were synthesized *in vitro* using T7 RNA Polymerase and DIG-RNA Labelling Mix (Roche Diagnostics); probes were then treated with RNase-free DNAse I according to the manufacturer's recommendations. Non-radioactive ISH was performed on 8 µm thin sections of *O. exaltata* and *O. sphegodes* labellum (two flowers/plant individuals per species) using 50 ng of probe as described previously [S10, 11]. *In situ* slides were photographed using a Leica DM6000B microscope (Leica Microsystems) and photos were merged using *Photomerge* in Adobe Photoshop CS5 (Adobe Systems).

Agro-infiltration, protoplast isolation and subcellular protein localization

Agrobacterium containing pMDC83-SAD-GFP constructs were cultured overnight at 28°C until OD_{600} = 1 was reached and resuspended in inoculation medium (10 mM $MgCl_2$, 10 mM 2-morpholinoethanesulphonic acid [MES], 100 µM acetosyringone) for 2 h at 4°C with gentle shaking. *Agrobacterium* containing pMDC84-PPH-GFP encoding a chloroplast-localized pheophytin pheophorbide hydrolase (PPH) [S5] and an untransformed culture were used as positive and negative controls, respectively. For transient expression of SAD-GFP fusion proteins, four young leaves of *Nicotiana benthamiana* (4 weeks old) were infiltrated with *Agrobacterium* using a plastic syringe. The plants were maintained in a growth chamber at 24°C under 16 h light. After 48 h, protoplasts were isolated from the infiltrated leaves using a previously described protocol [S12] with some modifications: (i) overlaying the protoplasts with 35% Percoll in MCP buffer (0.5 M sorbitol, 1 mM $CaCl_2$, 20 mM MES, pH 5.8) and (ii) storing the recovered protoplasts for 30 min at 4°C in the dark before visualization. Subcellular localization of *Ophrys* SADs was assessed by confocal laser-scanning microscopy of transiently expressed SAD-GFP fusion proteins in *N. benthamiana* leaf protoplasts. GFP fluorescence signals were induced with an excitation wavelength of 488 nm, and the emission signal was recovered between 495 and 530 nm. The signals were imaged using confocal laser-scanning microscopy (TCS SP2 & TCS SP5; Leica Microsystems) with 63× objectives. GFP and chlorophyll images were overlaid and autotuned using Adobe Photoshop CS5 (Adobe Systems).

SAD protein expression, purification and functional assay

The bacteria containing pET9d-mSAD constructs were grown in LB media containing ampicillin (100 µg/ml) and chloramphenicol (35 µg/ml) for 16 h at 37°C and protein expression was induced by supplementation with IPTG (1 mM) for 3 h at 37°C. Protein expression was assessed by Western blotting: protein extracts were separated on 12% SDS-PAGE, transferred to a PVDF membrane (Immobilon-P Membrane, Merck Millipore), detected by Anti-His (C-Term)-AP Antibody (Invitrogen) and visualized using NBT/BCIP tablets (Roche) and the manufacturers' protocols.

Proteins for use in enzyme assays were extracted by cation exchange chromatography where possible, or via their hexahistidine tag otherwise: SAD3 protein was extracted and purified by 20CM cation exchange chromatography (Perceptive Biosystems) as described previously [S3, 14]; since this method did not give satisfactory results with SAD4 and SAD5, Ni-NTA resin (Qiagen) was used for purification of these proteins [S15]. Soluble expression and purified protein could not be achieved for SAD1-A and SAD6 so that it was not possible to test these proteins for desaturase activity *in vitro*.

Biochemical desaturase activity assays were performed with recombinant SAD3, SAD4 and SAD5 proteins by gas chromatographic and mass spectrometric (GC/MS) analysis of reaction products from desaturase assays using the mass-labelled substrates 7,7,8,8-$^2H_4$-16:0-ACP (synthesized by A.P. Tulloch [S17] and kindly

provided by Patrick Covello, National Research Council Canada, Saskatoon, Canada) and 18,18,18-$^{2}H_3$-18:0-ACP (Cambridge Isotope Labs, Andover MA, USA) to distinguish between FAMEs of reaction products and background fatty acid contamination. Deuterated FAs were pure as determined by GC/MS of their FAMEs. Double-bond positions were determined by GC/MS analysis of the dimethyl disulphide (DMDS) derivatives of reaction products [S18]. Similarly, GC/MS-based assays with mass-labelled substrates were carried out for the engineered ('mutant') SAD5 proteins with single, double or triple amino acid replacements relative to SAD5-A. With few exceptions (see Table S1), GC/MS-based activity assays were run in triplicate.

Protein evolutionary analysis

Selected *SAD* coding sequences (Supplemental Experimental Procedures: Table listing selected sequences) were aligned in `MUSCLE` 3.8 [S19] based upon amino acid sequence, with minor manual adjustments in `BioEdit` 7.1.11 [S20] to take into account known protein structural features. Consensus sequences were produced in `BioEdit` for each *Ophrys* desaturase locus; separate allele-group specific consensus sequences were produced for *SAD1* and *SAD5* because these genes harboured several clear amino acid differences between distinct allele groups. Consensus sequences, the *Arabidopsis thaliana SSI2* and *Ricinus communis SAD* (*RcSAD*) outgroup sequences, and the known tree topology of their relationships [S1] were used for codon-based empirical Bayes reconstruction of ancestral sequences [S21] using the `baseml` and `codeml` programs of `PAML` 4.6 [S22]. Ancestral sequences based on different reconstructions (joint and marginal reconstructions using both `baseml` and `codeml` model M0) were very similar, with all sites of potential functional interest (including the substrate-binding cavity) reconstructed consistently. Reconstructions based on additional models (`codeml` 4.8 models M0 joint and M0, M1, M2, M7 and M8 marginal reconstructions) largely supported this inference; the `baseml`-based joint reconstructions were arbitrarily chosen for further analysis.

Based upon published data on protein structure and function, regions of particular molecular function were defined for desaturase proteins, as follows (except for transit peptide, amino acids refer to mature RcSAD protein):

- Transit peptide, defined as the first 33 amino acids of the unprocessed RcSAD protein (directly corresponding to the coding sequence) based on experimental data [S23];
- Secondary structural elements (α-helices and β-sheets), based on RcSAD crystallographic data [S24]: `KIEIFKSLDNWAEENI` (residues 36-51), `GFDEQVRELRERAK` (residues 75-88), `PDDYFVVLVGDMITEE` (residues 91-106), `LPTYQTMLNTL` (residues 108-118), `TSWAIWTRAWTAEENRHGDLLNKYLYLSG` (residues 130-158), `DMRQIEKTIQYLIGSG` (residues 161-176), `SPYLGFIYTSFQERATFISHGNTARQAKEH` (residues 184-213), `DIKLAQICGTIAADEKRHETAYTKIVEKLFE` (residues 215-245), `DGTVLAFADMMRKK` (residues 249-262), `LFDHFSAVAQRLG` (residues 278-290), `TAKDYADILEFLVGRWK` (residues 293-309), `SAEGQKAQDYVCRLPPRIRRLEERA` (residues 317-341), `PTMPF` (residues 349-353), `RQVKL` (residues 359-363);
- Salt bridges, based on the RcSAD crystal structure [S24]: E23, H30, K36, E82, R86, D92, D93, E105, D119, R122, D123, R137, E142, R145, R159, E182, R197, H203, D228, R231, E233, E241, K242, E245, D247, D257, R260, K261, H268, R274, D275, H281, R288, D299, R307, D325, R329, R335, R336, E339, D358, R359;
- Active site, based on the RcSAD crystal structure [S24]: E105, E143, H146, E196, E229, H232;
- Putative oxygen binding site, based on the RcSAD crystal structure [S24]: W139, T199;
- Putative electron transfer route, defined as the set of amino acids (out of two hypothetical sets based on the RcSAD crystal structure) [S24] that received experimental support [S25]: W62, H146, D228;
- Ferredoxin binding site, as supported by experimental data [S25]: K56, K60, K230;
- Desaturase protein dimerization interface, defined as residues of one polypeptide chain within 4.0 Å of the other polypeptide chain in the same dimer, based on RcSAD crystal structure [S24], using `RasMol` 2.7.5.2 software [S26]: F18, Q27, T29, H30, V58, E59, C61, Q63, P64, Q65, D66, L68, P69, P71, A72, F76, R84, E106, P109, Q112, T113, N116, T117, E124, T125, G126, A127, S128, R137, T140, A141, E142, N144, R145, D148, N151, K152, Y155, L156, G158, M162, R163, E166, K167, I169, Q170, Y171, I173, G174, S175, G176, M177, D178, P179, N183, G273, F357;
- ACP interacting sites, based on the crystal structure of the *Ricinus communis* SAD/stearoyl-ACP protein complex [S4]: R260, K262, D280, R333, R336, L337, R340, R344;
- Substrate channel, based on crystallographic data [S24, 27, 28]: T104, Y111, M114, L115, T117, L118, W139, P179, T181, G188, Y191, T192, Q195, T199, M265, P266, F279, S283, Y292.

It should be noted that due to fact that the alignment quality of the protein N-terminus and the poor prediction results for the transit peptide, the above definition of the transit peptide is the least reliable among the definitions provided and results stemming from it were treated with the necessary caution.

For each region of interest, the number of nonsynonymous (N) and synonymous (S) site changes was assessed between all pairs of extant sequences, and additionally between *SAD* loci along the phylogeny (using the ancestral sequences). This was done using the `GetCdsAvgCodonNS` function of the `biOP` software library (https://sourceforge.net/projects/biop/), considering the most parsimonious path(s) of codon change between any pair of sequences and the average of all codon changes in the case of any ambiguous nucleotide (i.e., parameter settings `PairwiseCodonNSPriorityMode = nsParsimonyAllSites` and `AmbiguityCodonNSPriorityMode = nsAverage`). The number of N and S site changes for a region of interest was compared against the respective number of changes in the remainder of the protein using Fisher's exact test in `R` 2.15.2 [S29].

Ancestral character reconstruction for Ophrys *desaturase function and localization*

Ancestral character states were inferred for three characters, namely C_{16} enzymatic activity, C_{18} enzymatic activity and subcellular localization, by maximum parsimony reconstruction in `R 2.15.2` [S29] using the `ape` library 3.0-11 [S30]. Present-day states for these characters were assigned based upon all available data as summarized in the table below.

Table listing present-day character states for *Ophrys* desaturase function and localization. This table lists the character states for modern orchid SADs (shown in Fig. 3) that were used for ancestral character reconstruction. The enzymatic activity columns indicate the character state assumed ('Δ9', 'Δ4' or 'LOF'), as well as the data and publication upon which it is based (biochemical assay or population-based association data). Subcellular localization was categorized as uniform (diffuse GFP signal throughout the plastid) or punctate (speckled GFP signal similar to plastoglobule localization). The housekeeping desaturases *Arabidopsis* SSI2 and *Ricinus* SAD display Δ^9 desaturation activity on 18:0-ACP and 16:0-ACP substrates and are localized in the plastid [S31, 32, 33, 34, 35].

Protein variant	C_{16} enzymatic activity	C_{18} enzymatic activity	Subcellular localization
SAD1-A	Δ4, population [S1]	Δ9, population [S1]	Unknown
SAD1-B	LOF, assay [S3]	LOF, assay [S3]	Uniform
SAD2	Δ4, assay [S3]	Δ9, assay [S3]	Punctate
SAD3	Δ9, assay (this study)	Δ9, assay (this study)	Uniform
SAD4	LOF, assay (this study)	LOF, assay (this study)	Punctate
SAD5-A	Δ9, assay (this study)	Δ9/LOF, assay (this study)[a]	Punctate
SAD5-B	LOF, population [S1][b]	LOF, population [S1][b]	Punctate[d]
SAD6	Δ9, population [S1]	Δ9/LOF, population [S1][c]	Punctate[d]

[a] Residual 18:0-ACP Δ^9 desaturase activity was coded either as 'LOF' or 'Δ9' and analyses performed separately with different character states.
[b] Since this protein variant carries the amino acid M141 instead of V141 found in SAD5-A, LOF is also consistent with complete lack of activity observed in biochemical assay data for SAD5-A^{V141M} (Table S1).
[c] Same character coding as SAD5-A because in some populations, an allelic variant of SAD6 may perform a function similar to SAD5-A [S1].
[d] Untested prediction based on sequence similarity.

Homology modelling and structural investigations

Homology modelling was performed on selected desaturase sequences (Supplemental Experimental Procedures: Table listing selected sequences) as in Schlüter et al. [S3], using the SWISS-MODEL server [S36] and an RcSAD crystal structure of 2.4 Å resolution [S37] (1OQ4 chain A) as the template. Homology models were superimposed on an RcSAD structure with 18:0 substrate modelled into it [S24], using `Swiss-PdbViewer` 4.1.0's "*Magic Fit*" function and all backbone atoms. Superimposed protein structures were then investigated in `RasMol` 2.7.5.2 software [S26]. In particular, to investigate the potential structural reason for activity differences among the desaturases SAD2, SAD3 and SAD5, amino acids differing between these proteins but within 6 Å of the modelled substrate were identified; to gauge the bottom of the substrate binding cavity, amino acids within 8 Å of the carbon atom at the substrate's aliphatic end were visualized.

Supplemental References

S1. Xu, S., Schlüter, P.M., Grossniklaus, U., and Schiestl, F.P. (2012). The genetic basis of pollinator adaptation in a sexually deceptive orchid. PLoS Genet. *8*, e1002889.

S2. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In The proteomics protocols handbook, J.M. Walker, ed. (Humana Press), pp. 571-607.

S3. Schlüter, P.M., Xu, S., Gagliardini, V., Whittle, E.J., Shanklin, J., Grossniklaus, U., and Schiestl, F.P. (2011). Stearoyl-acyl carrier protein desaturases are associated with floral isolation in sexually deceptive orchids. Proc. Natl. Acad. Sci. USA *108*, 5696-5701.

S4. Guy, J.E., Whittle, E.J., Moche, M., Lengqvist, J., Lindqvist, Y., and Shanklin, J. (2011). Remote control of regioselectivity in acyl-acyl carrier protein-desaturases. Proc. Natl. Acad. Sci. USA *108*, 16594-16599.

S5. Schelbert, S., Aubry, S., Burla, B., Agne, B., Kessler, F., Krupinska, K., and Hörtensteiner, S. (2009). Pheophytin pheophorbide hydrolase (pheophytinase) is involved in chlorophyll breakdown during leaf senescence in *Arabidopsis*. Plant Cell *21*, 767-785.

S6. Eugeni Piller, L., Besagni, C., Ksas, B., Rumeau, D., Bréhélin, C., Glauser, G., Kessler, F., and Havaux, M. (2011). Chloroplast lipid droplet type II NAD(P)H quinone oxidoreductase is essential for prenylquinone metabolism and vitamin K_1 accumulation. Proc. Natl. Acad. Sci. USA *108*, 14354-14359.

S7. Shumskaya, M., Bradbury, L.M.T., Monaco, R.R., and Wurtzel, E.T. (2012). Plastid localization of the key carotenoid enzyme phytoene synthase is altered by isozyme, allelic variation, and activity. Plant Cell *24*, 3725-3741.

S8. Lau, P.C.K., and Spencer, J.H. (1982). An efficient synthetic primer for the M13 cloning dideoxy sequencing system. Biosci. Rep. *2*, 687-696.

S9. Curtis, M.D., and Grossniklaus, U. (2003). A Gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol. *133*, 462-469.

S10. Sieber, P., Gheyselinck, J., Gross-Hardt, R., Laux, T., Grossniklaus, U., and Schneitz, K. (2004). Pattern formation during early ovule development in *Arabidopsis thaliana* Dev. Biol. *273*, 321-334.

S11. Karlgren, A., Carlsson, J., Gyllenstrand, N., Lagercrantz, U., and Sundström, J.F. (2009). Non-radioactive *in situ* hybridization protocol applicable for Norway spruce and a range of plant species. J. Vis. Exp. *26*, 1205.

S12. Endler, A., Meyer, S., Schelbert, S., Schneider, T., Weschke, W., Peters, S.W., Keller, F., Baginsky, S., Martinoia, E., and Schmidt, U.G. (2006). Identification of a vacuolar sucrose transporter in barley and *Arabidopsis* mesophyll cells by a tonoplast proteomic approach. Plant Physiol. *141*, 196-207.

S13. Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Prot. Sci. *8*, 978-984.

S14. Whittle, E.J., Tremblay, A.E., Buist, P.H., and Shanklin, J. (2008). Revealing the catalytic potential of an acyl-ACP desaturase: tandem selective oxidation of saturated fatty acids. Proc. Natl. Acad. Sci. USA *105*, 14738-14743.

S15. Andre, C., Kim, S.W., Yu, X.-H., and Shanklin, J. (2013). Fusing catalase to an alkane-producing enzyme maintains enzymatic activity by converting the inhibitory byproduct H_2O_2 to the cosubstrate O_2. Proc. Natl. Acad. Sci. USA *110*, 3191-3196.

S16. Cahoon, E.B., Cranmer, A.M., Shanklin, J., and Ohlrogge, J.B. (1994). Δ^6 hexadecenoic acid is synthesized by the activity of a soluble Δ^6 palmitoyl-acyl carrier protein desaturase in *Thunbergia alata* endosperm. J. Biol. Chem. *269*, 27519-27526.

S17. Tulloch, A.P. (1983). Synthesis, analysis and application of specifically deuterated lipids. Prog. Lipid Res. *22*, 235-256.

S18. Francis, G.W. (1981). Alkylthiolation for the determination of double-bond position in unsaturated fatty acid esters. Chem. Phys. Lipids *29*, 369-374.

S19. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. *32*, 1792-1797.

S20. Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. *41*, 95-98.

S21. Yang, Z., Kumar, S., and Nei, M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. Genetics *141*, 1641-1650.

S22. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586-1591.

S23. Thompson, G.A., Scherer, D.E., Foxall-van Aken, S., Kenny, J.W., Young, H.L., Shintani, D.K., Kridl, J.C., and Knauf, V.C. (1991). Primary structures of the precursor and mature forms of stearoyl-acyl

carrier protein desaturase from safflower embryos and requirement of ferredoxin for enzyme activity. Proc. Natl. Acad. Sci. USA *88*, 2578-2582.

S24. Lindqvist, Y., Huang, W., Schneider, G., and Shanklin, J. (1996). Crystal structure of Δ^9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. EMBO J. *15*, 4081-4092.

S25. Sobrado, P., Lyle, K.S., Kaul, S.P., Turco, M.M., Arabshahi, I., Marwah, A., and Fox, B.G. (2006). Identification of the binding region of the [2Fe-2S] ferredoxin in stearoyl-acyl carrier protein desaturase: insight into the catalytic complex and mechanism of action. Biochem. *45*, 4848-4858.

S26. Sayle, R.A., and Milnerwhite, E.J. (1995). RASMOL: biomolecular graphics for all. Trends Biochem. Sci. *20*, 374-376.

S27. Moche, M. (2003). Crystallographic studies of enzymes involved in biosynthesis of fatty acids; β-ketoacyl acyl carrier protein synthase II and Δ9-stearoyl-ACP desaturase. Ph.D. thesis, Karolinska Institutet (Stockholm, Sweden).

S28. Guy, J.E., Whittle, E., Kumaran, D., Lindqvist, Y., and Shanklin, J. (2007). The crystal structure of the ivy Δ^4-16:0-ACP desaturase reveals structural details of the oxidized active site and potential determinants of regio-selectivity. J. Biol. Chem. *282*, 19863-19871.

S29. R Development Core Team (2012). R: A language and environment for statistical computing. Version 2.15.2. (Vienna, Austria: R Foundation for Statistical Computing).

S30. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. Bioinformatics *20*, 289-290.

S31. Kachroo, A., Shanklin, J., Whittle, E., Lapchyk, L., Hildebrand, D., and Kachroo, P. (2007). The *Arabidopsis* stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. Plant Mol. Biol. *63*, 257-271.

S32. Joyard, J., Ferro, M., Masselon, C., Seigneurin-Berny, D., Salvi, D., Garin, J., and Rolland, N. (2010). Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. Prog. Lipid Res. *49*, 128-158.

S33. Shanklin, J., DeWitt, N.D., and Flanagan, J.M. (1995). The stroma of higher plant plastids contain ClpP and ClpC, functional homologs of *Escherichia coli* ClpP and ClpA: an archetypal two-component ATP-dependent protease. Plant Cell *7*, 1713-1722.

S34. Cahoon, E.B., Lindqvist, Y., Schneider, G., and Shanklin, J. (1997). Redesign of soluble fatty acid desaturase from plants for altered substrate specificity and double bond position. Proc. Natl. Acad. Sci. USA *94*, 4872-4877.

S35. Shanklin, J., and Somerville, C.R. (1991). Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. Proc. Natl. Acad. Sci. USA *88*, 2510-2514.

S36. Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucl. Acids Res. *31*, 3381-3385.

S37. Moche, M., Shanklin, J., Ghoshal, A., and Lindqvist, Y. (2003). Azide and acetate complexes plus two iron-depleted crystal structures of the di-iron enzyme Δ9 stearoyl-acyl carrier protein desaturase. J. Biol. Chem. *278*, 25072-25080.